

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
14 July 2005 (14.07.2005)

PCT

(10) International Publication Number  
**WO 2005/062881 A2**

(51) International Patent Classification: **Not classified**

William, C. [US/US]; 2225 Lakeridge Road, Grapevine, TX 76051 (US).

(21) International Application Number:  
PCT/US2004/043092

(74) Agent: **PRATT, John, S.**; Kilpatrick Stockton, LLC, Suite 2800, 1100 Peachtree Street, Atlanta, GA 30309 (US).

(22) International Filing Date:  
24 December 2004 (24.12.2004)

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/532,504 24 December 2003 (24.12.2003) US  
60/565,371 26 April 2004 (26.04.2004) US  
60/592,098 28 July 2004 (28.07.2004) US

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicants (*for all designated States except US*): **TRANSGENRX, INC.** [US/US]; 1755 Wittington Place Drive, Suite 300, Dallas, TX 75234 (US). **THE BOARD OF SUPERVISORS OF LOUISIANA STATE UNIVERSITY AND AGRICULTURAL AND MECHANICAL COLLEGE** [US/US]; P.O. Box 25055, Baton Rouge, LA 70894 (US).

**Published:**

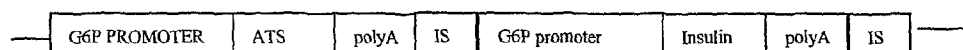
— *without international search report and to be republished upon receipt of that report*

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **COOPER, Richard, K.** [US/US]; 622 Woodview Court, Baton Rouge, LA 70810 (US). **ENRIGHT, Frederick, M.** [US/US]; 535 Woodleigh, Baton Rouge, LA 70810 (US). **FIORETTI,**

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: GENE THERAPY USING TRANSPOSON-BASED VECTORS



(57) Abstract: Methods and compositions are presented for the administration of transposon based vectors to an animal or human to provide gene therapy to the animal or human.



WO 2005/062881 A2

## 5 GENE THERAPY USING TRANSPOSON-BASED VECTORS

## FIELD OF THE INVENTION

The present invention relates generally to use of transposon-based vectors in the preparation of a medicament useful for providing therapy, including gene therapy,  
10 to animals and humans following administration of the medicament. The vectors of the present invention may be directed to specific tissues, organs and cells where a selected gene is stably incorporated and produces proteins, peptides or nucleic acids which have a therapeutic effect in the animal or human.

## 15 BACKGROUND OF THE INVENTION

Improved gene delivery technologies are needed for the treatment of disease in animals. Many diseases and conditions can be treated with gene-delivery technologies, which provide a gene of interest to an animal suffering from the disease or the condition. An example of such disease is Type 1 diabetes. Type 1 diabetes is  
20 an autoimmune disease that ultimately results in destruction of the insulin producing  $\beta$ -cells in the pancreas. Although animals with Type 1 diabetes may be treated adequately with insulin injections or insulin pumps, these therapies are only partially effective. In addition, hyper- and hypoglycemia occurs frequently despite intensive home blood glucose monitoring. Finally, careful dietary constraints are needed to  
25 maintain an adequate ratio of calories consumed. Development of gene therapies providing delivery of the insulin gene into the pancreas of diabetic animals could overcome many of these problems and result in improved life expectancy and quality of life.

Several of the prior art gene delivery technologies employed viruses that are  
30 associated with potentially undesirable side effects and safety concerns. The majority of current gene-delivery technologies useful for gene therapy rely on virus-based delivery vectors, such as adeno and adeno-associated viruses, retroviruses, and other

viruses, which have been attenuated to no longer replicate. (Kay, M.A., et al. 2001. Nature Medicine 7:33-40).

There are multiple problems associated with the use of viral vectors. First, they are not tissue-specific. In fact, a gene therapy trial using adenovirus was recently  
5 halted because the vector was present in a patient's sperm (Gene trial to proceed despite fears that therapy could change child's genetic makeup. The New York Times, December 23, 2001). Second, viral vectors are likely to be transiently incorporated, which necessitates re-treating a patient at specified time intervals. (Kay, M.A., et al. 2001. Nature Medicine 7:33-40). Third, there is a concern that a viral-  
10 based vector could revert to its virulent form and cause disease. Fourth, viral-based vectors require a dividing cell for stable integration. Fifth, viral-based vectors indiscriminately integrate into various cells, which can result in undesirable germline integration. Sixth, the required high titers needed to achieve the desired effect have resulted in the death of one patient and they are believed to be responsible for  
15 induction of cancer in a separate study. (Science, News of the Week, October 4, 2002).

Accordingly, what is needed is a new method to produce transgenic animals and humans with stably incorporated genes, in which the vector containing those genes does not cause disease or other unwanted side effects. There is also a need for  
20 DNA constructs that would be stably incorporated into the tissues and cells of animals and humans, including cells in the resting state that are not replicating. There is a further recognized need in the art for DNA constructs capable of delivering genes to specific tissues and cells of animals and humans and for producing proteins in those animals and humans.

25 When incorporating a gene of interest into an animal or human for the production of a desired protein or when incorporating a gene of interest in an animal for the treatment of a disease, it is often desirable to selectively activate incorporated genes using inducible promoters. These inducible promoters are regulated by substances either produced or recognized by the transcription control elements within  
30 the cell in which the gene is incorporated. In many instances, control of gene expression is desired in transgenic animals and humans so that incorporated genes are selectively activated at desired times and/or under the influence of specific substances. Accordingly, what is needed is a means to selectively activate genes introduced into the genome of cells of a transgenic animal or human. This can be

taken a step further to cause incorporation to be cell-specific , which prevents widespread gene incorporation throughout the body. This decreases the amount of DNA needed for a treatment, decreases the chance of incorporation in gametes, and targets gene delivery, incorporation, and expression to the desired tissue where the gene is needed to function.

RNAi has been targeted as a tool for several uses including treatment of genetic abnormalities and disease, cancer, and development. There are mainly two types of short RNAs that target complementary messengers in animals: small interfering RNAs and micro-RNAs. Both are produced by the cleavage of double-stranded RNA precursors by Dicer, a member of the RNase III family of double-stranded specific endonucleases, and both guide the RNA-induced silencing complex to cleave specifically RNAs sharing sequence identity with them. RNAi technology can be used in therapeutic approaches to treat disease and various conditions. However, a major drawback to RNAi therapy has been the lack of a reliable delivery method of the short RNA sequences. Most researchers working in the field rely on producing short double stranded RNA (dsRNA) in the laboratory and then delivering these short dsRNAs either by direct injection, electroporation, by complexing with a transfecting reagent, etc. The result is gene silencing, but only as long as the dsRNA remains present in the cell, which generally begins to decrease after about 20 h. In order to obtain lasting therapeutic effects, the RNAi sequence must be expressed long term, preferably under a constitutive promoter. In order to accomplish RNAi expression in a plasmid-based vector and subsequent recognition by RNA induced silencing complex (RISC), the RNA must be double stranded. To obtain dsRNA from a vector, it must be expressed as a short hairpin RNA (shRNA), in which there is a sense strand, a hairpin loop region and an antisense strand (M. Izquierdo. 2004. *Short interfering RNAs as a tool for cancer gene therapy*. *Cancer Gene Therapy* pp 1-11; Miyagishi et al. 2004. *J Gene Med* 6:715-723). The hairpin region allows the antisense strand to loop back and bind to the complimentary sense strand.

## SUMMARY OF THE INVENTION

The present invention addresses the problems described above by providing new, effective and efficient compositions comprising transposon-based vectors for providing therapy, including gene therapy, to animals and humans. The present invention provides methods of using these compositions for providing therapy to

animals and humans. These transposon-based vectors are used in the preparation of a medicament useful for providing a desired effect to a recipient following administration. Gene therapy includes, but is not limited to, introduction of a gene into an animal using a transposon-based vector. Such genes are called exogenous  
5 genes although it is to be understood that these genes may also be found in the recipient animal. These genes may serve a variety of functions in the recipient such as coding for the production of nucleic acids, for example RNA, or coding for the production of proteins and peptides. An advantage of the present invention is that transgenic animals are produced with higher efficiencies than observed in the prior  
10 art, including efficient incorporation of large polynucleotide sequences. The present invention facilitates efficient incorporation of the polynucleotide sequences, including the genes of interest, promoters, insertion sequences and poly A, with transfection efficiencies of at least 30%. Transfection efficiencies greater than 30%, 40%, 50%, 60% and 70% have been observed.

15 Transgenic animals further include but are not limited to avians, fish, amphibians, reptiles, insects, and mammals. It is to be understood that humans are encompassed within the term "animal" and the term "mammal" in the present application. In another embodiment, the animal is a milk-producing animal, including but not limited to bovine, porcine, ovine and equine animals. Transgenic animals  
20 include all egg-laying animals and milk-producing animals. In one embodiment, the animal is an avian animal. In another embodiment, the animal is a mammal. Preferred animals may be pets, domestic animals, exotic animals, zoo animals, wild animals or any other type of animal in need of gene therapy. Animals are made *transgenic through administration of a composition comprising a transposon-based*  
25 *vector designed for stable incorporation of a gene of interest for production of a desired protein, peptide or nucleic acid, together with an acceptable carrier.*

The present invention addresses the problems described above by providing methods and compositions comprising transposon-based vectors for stable incorporation of an exogenous gene into a specific cell or tissue of an animal. It is to  
30 be understood that the terms cell-specific and tissue-specific are often used interchangeably by those of ordinary skill in the art. In the present application, cell-specific promoters indicate a promoter that is active in that cell. A promoter may be used to drive the transposase that may or may not drive the desired target protein. For example, a vitellogenin promoter or a glucose-6-phosphatase promoter (both

promoters being found in hepatocytes) may be used to direct incorporation of the gene into the liver, but the promoter driving the target gene/protein, may be constitutive, cell-specific, or be induced by a hormone, antibiotic, etc., that is not specific to a hepatocyte.

5 In one embodiment of the present invention, the transposon-based vectors are designed for cell-specific gene expression, for example, by placing a selected gene under control of a cell-specific promoter, which further increases cell or tissue specificity of expression of the selected gene. In one embodiment of the present invention, the vectors are used for gene therapy of animals or humans, wherein the  
10 expression of the selected gene has a therapeutic effect on a cell or a tissue in which it is specifically incorporated, expressed, or both. The expression of the selected gene may also have a therapeutic effect on other cells or tissues, particularly when the expressed protein is secreted from the cell and can access other cells. In another embodiment, the vectors are used to integrate a desired gene into specific cells of an  
15 animal for production of biologic agents encoded by the gene of interest.

This invention provides polynucleotide cassettes or vectors containing at least one gene of interest and at least one pro polynucleotide sequences, wherein the at least one gene of interest is operably-linked to a pro nucleotide sequence. Each of the at least one gene of interest encodes a polypeptide. This invention also provides  
20 polynucleotide cassettes containing two or more genes of interest and two or more pro polynucleotide sequences, wherein each gene of interest is operably-linked to a pro nucleotide sequence. Each of the genes of interest encodes a polypeptide that forms a part of the multimeric protein. One discovery of the present invention is the use of  
25 pro portions of prepro signal sequences to facilitate appropriate processing, expression, and/or formation of multimeric proteins in an individual. Several examples of prepro polynucleotides from which a pro polynucleotide can be derived or be a part of are a cecropin prepro, lysozyme prepro, ovomucin prepro, ovotransferrin prepro, a signal peptide for tumor necrosis factor receptor (SEQ ID NO:1). Signal sequences for protein secretion are readily available through databases  
30 such as GenBank or the literature. Signal sequences for protein secretion can be identified by one of ordinary skill in the art, for example through comparison of mRNA to mature protein and identification of the sequence removed from the secreted protein. The prepro or pro polynucleotide can be a cecropin prepro or pro polynucleotide selected from the group consisting of cecropin A1, cecropin A2,

cecropin B, cecropin C, cecropin D, cecropin E and cecropin F. In a preferred embodiment, the pro polynucleotide is a cecropin B pro polynucleotide having a sequence shown in SEQ ID NO:2 or SEQ ID NO:3. A preferred prepro polynucleotide is a cecropin B polynucleotide having a sequence shown in SEQ ID NO:4 or SEQ ID NO:5.

Another discovery of the present invention is that cecropin prepro sequences facilitate appropriate processing, expression, and/or formation of proteins, including multimeric proteins, in an individual. Accordingly, the present invention includes polynucleotide cassettes containing one or more genes of interest operably-linked to a cecropin prepro sequence. In one embodiment, the polynucleotide cassette contains two or more genes of interest operably-linked to a cecropin prepro sequence. Preferred cecropin prepro polynucleotides are provided in SEQ ID NO:4 and SEQ ID NO:5. The present invention also includes polynucleotide cassettes containing two or more genes of interest operably linked to a cecropin prepro polynucleotide, wherein pro sequences are located between the genes of interest.

These polynucleotide cassettes are administered to an individual for expression of polypeptide sequences and the formation of a protein or a multimeric protein.

In another embodiment of the present invention, the gene of interest incorporated into the cell is expressed and the resulting protein or peptide has regulatory properties affecting a function of the cell or tissue in which it is expressed.

In a further embodiment, the gene of interest incorporated into the cell is expressed, secreted and affects another cell.

In another embodiment of the present invention, the gene of interest incorporated into the cell is expressed as an inhibitory molecule, such as an RNAi may regulate the expression or overexpression of another substance.

Cell or tissue-specific expression of a gene of interest may be particularly advantageous because of the possible toxic or otherwise potentially negative effects of the gene of interest if it were expressed in an undesirable location. Control of expression in the desired cell or tissue is achieved by operably linking the gene of interest to a cell or tissue specific regulator.

In yet another embodiment of the present invention, the expression of the exogenous gene is blocked unless an animal is at a selected stage in its life cycle. For example, expression of an exogenous gene may not be desirable until an animal

reaches puberty. In this embodiment, the expression of the exogenous gene is controlled by a promoter, which is activated specifically at the desired stage in the life of the animal. Alternatively, expression of an exogenous gene may not be desirable until a disease process begins or at a time when a substance, such as a hormone or an inflammatory molecule, is found in undesirable levels.

In yet another embodiment of the present invention, the expression of the exogenous gene is blocked unless an animal produces a substance. For example, expression of an exogenous gene may not be desirable until an animal begins producing cancer-related molecules. In this embodiment, the expression of the exogenous gene is controlled by a promoter, which is activated specifically by the cancer-related molecules. Such gene therapy fights the disease process at very early stages.

An additional advantage of the present invention is that a disease or a condition can be treated in a specific organ or tissue of an animal without risk of making other organs or tissues transgenic. This is particularly useful when concerns exist about passing a transgene to the progeny of the transgenic animal, or contaminating the environment with the transgene shed by the animal. The methods and compositions of the present invention are particularly advantageous in applications where germline integration of exogenous DNA is undesirable.

The compositions of the present invention may be introduced into an animal through any route of administration that serves to deliver the composition to the desired organs, tissues and cells. Such routes of administration include, but are not limited to, oral and parenteral routes such as intravascular, intravenous, intraarterial, intracardiac, intraperitoneal, intramuscular, anal, intracerebrovascular, intracerebroventricular, cutaneous, intradermal, subcutaneous, transdermal, into any duct system, into any cavity or space, such as the abdominopelvic, pleural, pericardial, peritoneal cavities or spaces, intrathecal, or into the respiratory system, the urinary system, the gastrointestinal system, the nervous system, the lymphatic system, the immune system, the reproductive system and the endocrine system.

Administration of the transposon based vectors into the cardiovascular system achieves rapid distribution throughout the animal to reach target tissues and cells receiving blood supply. Such administration may be into any chamber of the heart, for example into the left ventricle, the right ventricle, or the atrial chambers, for rapid distribution into the systemic circulation or the pulmonary circulation. The vectors



may be administered into selected vessels, such as the cardiac vessels, into the aorta or into a selected vessel leading to a targeted group of cells, a tissue, an organ or a tumor. Administration into the left side of the heart may target the systemic circulation through the aorta and any of its branches, including but not limited to the coronary vessels, the ovarian or testicular arteries, the renal arteries, the arteries  
5 supplying the gastrointestinal and pelvic tissues, including the celiac, cranial mesenteric and caudal mesenteric vessels and their branches, the common iliac arteries and their branches to the pelvic organs, the gastrointestinal system and the lower extremity, the carotid, brachiocephalic and subclavian arteries. It is to be  
10 understood that the specific names of blood vessels change with the species under consideration and are known to one of ordinary skill in the art. Administration into the left ventricle or ascending aorta supplies any of the tissues receiving blood supply from the aorta and its branches, including but not limited to the testes, ovary, oviduct, and liver. Administration may occur through any means, for example by injection  
15 into the left ventricle, or by administration through a cannula or needle introduced into the left atrium, left ventricle, aorta or a branch thereof.

The compositions of the present invention may be administered to a reproductive organ including, but not limited to, an oviduct, an ovary, the testes, seminal vesicle, any accessory organ, or into the duct system of the mammary gland.  
20 The compositions of the present invention may be administered to a reproductive organ of an animal through the cloaca. The compositions of the present invention may be directly administered to an organ or can be administered to an artery or vein leading to the organ. A transfection reagent is optionally added to the composition before administration.

25 The transposon-based vectors of the present invention include a transposase, operably-linked to a first promoter, and a coding sequence for a protein or peptide of interest operably-linked to a second promoter, wherein the coding sequence for the protein or peptide of interest and its operably-linked promoter are flanked by transposase insertion sequences recognized by the transposase. The transposon-based  
30 vector also includes the following characteristics: a) one or more modified Kozak sequences at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the nucleotide at the third base position of each codon is changed to an A or a T without changing the corresponding amino acid; c) addition of

one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. In some embodiments, the effective polyA sequence is an avian optimized polyA sequence.

5           The present invention also provides for tissue-specific incorporation and/or expression of a gene of interest. Tissue-specific stable incorporation of a gene of interest may be achieved by placing the transposase gene under the control of a tissue-specific promoter, whereas tissue-specific expression of a gene of interest may be achieved by placing the gene of interest under the control of a tissue-specific  
10 promoter. Such tissues include all tissues within the body, for example, connective tissue, muscle, bone, lymphoid tissue, and nervous tissue. In some embodiments, the gene of interest is transcribed under the influence of an ovalbumin, or other oviduct specific, promoter. In other embodiments, promoters may be specific for cells in another organ including but not limited to the liver, brain, mammary gland, any  
15 endocrine organ, and thymus.

          The present invention also provides for cell-specific incorporation and/or expression of a gene of interest. Cell-specific incorporation of a gene of interest may be achieved by placing the transposase gene under the control of a cell-specific promoter, whereas tissue-specific expression of a gene of interest may be achieved by  
20 placing the gene of interest under the control of a cell-specific promoter. Such promoters may include promoters specific for cells such as neurons, glia, hepatocytes, epithelial cells, cells of the immune system, fibroblasts, chondrocytes, synovial cells, osteoblasts, osteocytes, osteoclasts, muscle cells (including striated, smooth and cardiac muscle cells), granulocytes, lymphocytes, T lymphocytes, B-lymphocytes,  
25 thymocytes, germ cells, blast cells, cancerous cells, endocrine cells, white blood cells, pancreatic islet cells, acinar cells, splenocytes, follicular cells, and so on. Cell specific promoters are known to one of ordinary skill in the art.

          The present invention advantageously produces a high number of transgenic animals having a gene of interest stably incorporated. In some embodiments wherein  
30 the transposon-based vector is administered to the ovary or the testes, these transgenic animals successfully pass the desired gene to their progeny. Accordingly, the present invention can be used to obtain transgenic animals having the gene of interest incorporated into the germline through transfection of the ovary or testes. These transgenic animals of the present invention produce large amounts of a desired

molecule encoded by the transgene. Such germline transmission can produce generations of animals containing the desired gene. Such a gene may be useful in providing gene therapy to animals known to be susceptible to specific conditions, for example an immune deficiency or arthritis.

5           Gene therapy of cells other than germline cells may also be achieved by introducing the transposon-based vectors to these cells for stable incorporation of exogenous genes.

Any desired gene may be incorporated into the novel transposon-based vectors of the present invention in order to synthesize a desired molecule in the transgenic  
10 animals to provide gene therapy. Proteins, peptides and nucleic acids are preferred desired molecules to be produced by the transgenic animals of the present invention. In order to provide gene therapy to an animal, the gene desired to produce a desired molecule in the transgenic animal is selected and inserted into the transposon-based vectors of the present invention.

15           Nucleic acids may be made by the transgenic animals. Such nucleic acids include, but are not limited to, single stranded DNA, RNA, antisense nucleic acids, siRNA, and polynucleotide strands that affect cellular function. Some of these nucleic acids may affect cellular function by modulating transcription of a gene, for example by inhibiting the function of the gene which may be producing undesirable  
20 effects such as inappropriate amounts of molecules or molecules that may be deleterious to the animal.

Genes may be regulated by regulating a specific transcription factor. Cystic fibrosis, for example, is the result of a mutant protein. Even if RNAi is used, the cell is still synthesizing mutant RNA. However, if the transcription factor allowing  
25 expression of the mutant protein is blocked, then the mutant gene is shut down and the normal gene can be expressed under a different promoter.

A wide range of recombinant peptides and proteins can be produced in animals receiving the gene therapy of the present invention. Enzymes, hormones, antibodies, growth factors, serum proteins, commodity proteins, fusion proteins,  
30 fusion peptides, biological response modifiers, cytokines, chemoattractants, chemorepellents, receptor agonists, receptor antagonists, peptides and designed proteins may all be made through practice of the present invention.

Accordingly, it is an object of the present invention to provide novel transposon-based vectors useful in providing gene therapy to an animal.

It is an object of the present invention to provide novel transposon-based vectors for use in the preparation of a medicament useful in providing gene therapy to an animal or human.

It is another object of the present invention to provide novel transposon-based  
5 vectors that encode for the production of desired proteins or peptides in cells.

Yet another object of the present invention to provide novel transposon-based vectors that encode for the production of desired nucleic acids in cells.

It is a further object of the present invention to provide methods for cell and tissue specific incorporation of transposon-based DNA constructs comprising  
10 targeting a selected gene to a specific cell or tissue of an animal.

It is yet another object of the present invention to provide methods for cell and tissue specific expression of transposon-based DNA constructs comprising designing a DNA construct with cell specific promoters that enhance stable incorporation of the selected gene by the transposase and expressing the selected gene in the cell.

15 It is an object of the present invention to provide gene therapy for generations through germ line administration of a transposon-based vector.

Another object of the present invention is to provide gene therapy in animals through non germ line administration of a transposon-based vector.

It is further an object of the present invention to provide a method to produce  
20 transgenic animals through intraovarian or intratesticular administration of a transposon-based vector that are capable of producing transgenic progeny.

It is further an object of the present invention to provide a method to produce transgenic animals through cardiovascular administration of a transposon-based vector.

25 Another object of the present invention is to provide gene therapy in animals through administration of a transposon-based vector, wherein the animals produce desired proteins, peptides or nucleic acids.

Yet another object of the present invention is to provide gene therapy in animals through administration of a transposon-based vector, wherein the animals  
30 produce desired proteins or peptides that are recognized by receptors on target cells.

Still another object of the present invention is to provide gene therapy in animals through administration of a transposon-based vector, wherein the animals produce desired fusion proteins or fusion peptides, a portion of which are recognized

by receptors on target cells, in order to deliver the other protein or peptide component of the fusion protein or fusion peptide to the cell to induce a biological response.

Yet another object of the present invention is to provide a method for gene therapy of animals through administration of transposon-based vectors comprising  
5 tissue specific promoters and a gene of interest to facilitate tissue specific incorporation and expression of a gene of interest to produce a desired protein, peptide or nucleic acid.

Another object of the present invention is to provide a method for gene therapy of animals through administration of transposon-based vectors comprising  
10 cell specific promoters and a gene of interest to facilitate cell specific incorporation and expression of a gene of interest to produce a desired protein, peptide or nucleic acid.

Still another object of the present invention is to provide a method for gene therapy of animals through administration of transposon-based vectors comprising  
15 cell specific promoters and a gene of interest to facilitate cell specific incorporation and expression of a gene of interest to produce a desired protein, peptide or nucleic acid, wherein the desired protein, peptide or nucleic acid has a desired biological effect in the animal.

Another object of the present invention is to provide transgenic animals that  
20 contain a stably incorporated transgene.

An advantage of the present invention is that transgenic animals are produced with higher efficiencies than observed in the prior art.

Another advantage of the present invention is that transgenic animals are produced with higher efficiencies than observed in the prior art, including large  
25 transgenes.

Another advantage of the present invention is that these transgenic animals possess high copy numbers of the transgene.

Another advantage of the present invention is that the transgenic animals produce large amounts of desired molecules encoded by the transgene.

30 Still another advantage of the present invention is that desired molecules are produced by the transgenic animals much more efficiently and economically than prior art methods, thereby providing a means for efficient gene therapy.

Yet another advantage of the present invention is that the desired proteins and peptides are produced rapidly after making animals transgenic through introduction of the vectors of the present invention.

These and other objects, features and advantages of the present invention will  
5 become apparent after a review of the following detailed description of the disclosed embodiments and claims.

#### BRIEF DESCRIPTION OF THE FIGURES

Figure 1 depicts schematically a transposon-based vector containing a  
10 transposase operably linked to a first promoter and a gene of interest and poly A operably-linked to a second promoter, wherein the gene of interest and its operably-linked promoter are flanked by insertion sequences (IS) recognized by the transposase. "Pro" designates a promoter. In this and subsequent figures, the size of the actual nucleotide sequence is not necessarily proportionate to the box representing  
15 that sequence.

Figure 2 depicts schematically a transposon-based vector for targeting deposition of a polypeptide in an egg white wherein Ov pro is the ovalbumin promoter, Ov protein is the ovalbumin protein and PolyA is a polyadenylation sequence. The TAG sequence includes a spacer sequence, the gp41 hairpin loop from  
20 HIV I and a protease cleavage site.

Figure 3 depicts schematically a transposon-based vector for targeting deposition of a polypeptide in an egg white wherein Ovo pro is the ovomucoid promoter and Ovo SS is the ovomucoid signal sequence. The TAG sequence includes a spacer, the gp41 hairpin loop from HIV I and a protease cleavage site.

Figure 4 depicts schematically a transposon based-vector for expression of an RNAi molecule. "Tet<sub>i</sub> pro" indicates a tetracycline inducible promoter whereas "pro" indicates the pro portion of a prepro sequence as described herein. "Ovgen" indicates approximately 60 base pairs of an ovalbumin gene, "Ovotrans" indicates approximately 60 base pairs of an ovotransferring gene and "Ovomucin" indicates  
30 approximately 60 base pairs of an ovomucin gene.

Figure 5 is a schematic of a vector for stable transformation of liver cells for production of insulin. This vector provides for stable transformation of hepatocytes and allows the production of insulin in response to blood glucose levels.

Figure 6 depicts schematically a transposon-based vector targeted to hepatocytes for use in gene therapy for diabetes. A glucose-6-phosphatase (G-6-P) promoter is placed upstream of the transposase (ATS) and another G-6-P promoter is upstream of the proinsulin gene and the polyadenylation sequence (PolyA). Insertion sequences (IS) recognized by the transposase flank the G-6-P promoter, proinsulin gene and PolyA.

Figure 7 depicts schematically a transposon-based vector targeted to hepatocytes for use in gene therapy for production of growth hormone and treatment of growth hormone deficiencies. A glucose-6-phosphatase (G-6-P) promoter is placed upstream of the transposase (ATS) and an albumin promoter is upstream of the human growth hormone (hGH) gene and the polyadenylation sequence (PolyA). Insertion sequences (IS) recognized by the transposase flank the albumin promoter, GH gene and PolyA.

Figure 8 depicts schematically a transposon-based vector targeted to respiratory epithelial cells for use in gene therapy for treatment of cystic fibrosis. A ciliated cell-specific promoter (FOXJ1) is placed upstream of the transposase (ATS), normal CFTR gene and the polyadenylation sequence (PolyA). Insertion sequences (IS) recognized by the transposase flank the normal cystic fibrosis transmembrane conductance regulator gene (CFTR) and PolyA.

Figure 9 depicts schematically a transposon-based vector targeted to cancer cells for use in gene therapy for treatment of cancer. Human telomerase reverse transcriptase/human surfactant protein A1 (HTRT/hSPA1) is placed upstream of the transposase (ATS). A SV40p promoter is placed upstream of the a gene for cholera toxin and the polyadenylation sequence (PolyA). Insertion sequences (IS) recognized by the transposase flank the SV40p promoter, cholera toxin gene and PolyA.

#### DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a new, effective and efficient method of providing gene therapy to animals through administration of a composition comprising a transposon-based vector designed for incorporation of a gene of interest and production of a desired molecule. These transposon-based vectors are used in the preparation of a medicament for administration to an animal or human to provide a beneficial effect in the recipient through production of a desired molecule. The present invention facilitates efficient incorporation of the polynucleotide sequences,

including the genes of interest, promoters, insertion sequences and poly A, with transfection efficiencies of at least 30%. Transfection efficiencies greater than 30%, 40%, 50%, 60% and 70% have been observed. Desired molecules encoded by genes of interest include, but are not limited to, nucleic acids, proteins and peptides.

5 Proteins include multimeric proteins. Multimeric proteins include associated multimeric proteins (two or more associated polypeptides) and multivalent multimeric proteins (a single polypeptide encoded by more than one gene of interest). Expression and/or formation of the multimeric protein in the individual is achieved by administering a polynucleotide cassette containing the genes of interest to the

10 individual. The polynucleotide cassette may additionally contain one or more pro sequences, prepro sequences, cecropin prepro sequences, and/or cleavage site sequences. In a preferred embodiment, the polynucleotide cassette is administered through the vascular system. Nucleic acids that may be produced in transfected cells include single stranded DNA, RNA, antisense nucleic acids, siRNA, and

15 polynucleotide strands that affect cellular function.

#### Definitions

It is to be understood that as used in the specification and in the claims, “a” or “an” can mean one or more, depending upon the context in which it is used. Thus, for

20 example, reference to “a cell” can mean that at least one cell can be utilized.

The term “animal” includes a human in the present application.

The term “protein” includes multimeric protein” as described in PCT US03/41261. Multimeric proteins include associated multimeric proteins (two or more associated polypeptides) and multivalent multimeric proteins (a single

25 polypeptide encoded by more than one gene of interest).

The term “nucleic acid” includes double stranded DNA, single stranded DNA, RNA, antisense nucleic acids, siRNA, and polynucleotide strands.

The term “antibody” is used interchangeably with the term “immunoglobulin” and is defined herein as a protein synthesized by an animal or a cell of the immune

30 system in response to the presence of a foreign substance commonly referred to as an “antigen” or an “immunogen”. The term antibody includes fragments of antibodies. Antibodies are characterized by specific affinity to a site on the antigen, wherein the site is referred to an “antigenic determinant” or an “epitope”. Antigens can be naturally occurring or artificially engineered. Artificially engineered antigens



include, but are not limited to, small molecules, such as small peptides, attached to haptens such as macromolecules, for example proteins, nucleic acids, or polysaccharides. Artificially designed or engineered variants of naturally occurring antibodies and artificially designed or engineered antibodies not occurring in nature  
5 are all included in the current definition. Such variants include conservatively substituted amino acids and other forms of substitution as described in the section concerning proteins and polypeptides.

The present invention provides novel transposon-based vectors and their use for specific and stable incorporation of a gene of interest into a specific cell to stably  
10 incorporate the gene.

The term "gene" is defined herein to include a coding region for a protein, peptide or polypeptide.

The term "transgenic animal" refers to an animal having at least a portion of the transposon-based vector DNA is incorporated into its DNA. While a transgenic  
15 animal includes an animal wherein the transposon-based vector DNA is incorporated into the germline DNA, a transgenic animal also includes an animal having DNA in one or more cells that contain a portion of the transposon-based vector DNA for any period of time. In a preferred embodiment, a portion of the transposon-based vector comprises a gene of interest. More preferably, the gene of interest is incorporated into  
20 the animal's DNA for a period of at least a few days, preferably the reproductive life of the animal, and preferably the life of the animal.

The term "vector" is used interchangeably with the terms "construct", "DNA construct", "genetic construct", and "polynucleotide cassette" to denote synthetic nucleotide sequences used for manipulation of genetic material, including but not  
25 limited to cloning, subcloning, sequencing, or introduction of exogenous genetic material into cells, tissues or organisms, such as animals. It is understood by one skilled in the art that vectors may contain synthetic DNA sequences, naturally occurring DNA sequences, or both. The vectors of the present invention are transposon-based vectors as described herein.

30 When referring to two nucleotide sequences, one being a regulatory sequence, the term "operably-linked" is defined herein to mean that the two sequences are associated in a manner that allows the regulatory sequence to affect expression of the other nucleotide sequence. It is not required that the operably-linked sequences be directly adjacent to one another with no intervening sequence(s).

The term "regulatory sequence" is defined herein as including promoters, enhancers and other expression control elements such as polyadenylation sequences, matrix attachment sites, insulator regions for expression of multiple genes on a single construct, ribosome entry/attachment sites, introns that are able to enhance  
5 expression, and silencers. Promoters may be cell specific or tissue specific to facilitate expression in a desired target.

#### Transposon-Based Vectors

While not wanting to be bound by the following statement, it is believed that  
10 the nature of the DNA construct is an important factor in successfully providing gene therapy to animals. The "standard" types of plasmid and viral vectors that have previously been almost universally used for transgenic work in all species have low efficiencies and may constitute a major reason for the low rates of transformation previously observed. The DNA (or RNA) constructs previously used often do not  
15 integrate into the host DNA, or integrate only at low frequencies. Other factors may have also played a part, such as poor entry of the vector into target cells. The present invention provides transposon-based vectors that can be administered to an animal that overcome the prior art problems relating to low transgene integration frequencies. In the present invention integration frequencies greater than 30%, 40%, 50%, 60%  
20 and also 70% are often obtained for vectors about 10 kb or less. In some cases, depending on the route of administration, integration frequencies of over 70%, 80% and even 90% are observed. If the vector is over 15 kb, then transfection rates of over 30% and over 40% are feasible. Two preferred transposon-based vectors of the present invention in which a transposase, gene of interest and other polynucleotide  
25 sequences may be introduced are termed pTnMCS (SEQ ID NO:6) and timed (SEQ ID NO:7).

The transposon-based vectors of the present invention produce integration frequencies an order of magnitude greater than has been achieved with previous vectors. More specifically, intratesticular injections performed with a prior art  
30 transposon-based vector (described in U.S. Patent No. 5,719,055) resulted in 41% sperm positive roosters whereas intratesticular injections performed with the novel transposon-based vectors of the present invention resulted in 77% sperm positive roosters. Actual frequencies of integration were estimated by either or both

comparative strength of the PCR signal from the sperm and histological evaluation of the testes and sperm by quantitative PCR.

The transposon-based vectors of the present invention include a transposase gene operably-linked to a first promoter, and a coding sequence for a desired protein or peptide operably-linked to a second promoter, wherein the coding sequence for the desired protein or peptide and its operably-linked promoter are flanked by transposase insertion sequences recognized by the transposase. The transposon-based vector also includes one or more of the following characteristics: a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:8) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. The transposon-based vector may additionally or alternatively include one or more of the following Kozak sequences at the 3' end of any promoter, including the promoter operably-linked to the transposase: ACCATGG (SEQ ID NO:9), AAGATGT (SEQ ID NO:10), ACGATGA (SEQ ID NO:11), AAGATGG (SEQ ID NO:12), GACATGA (SEQ ID NO:13), ACCATGA (SEQ ID NO:14), and ACCATGA (SEQ ID NO:15), ACCATGT (SEQ ID NO:16). In another embodiment, the transposon-based vector comprises an avian optimized polyA sequence and does not comprise a modified Kozak sequence.

Figure 1 shows a schematic representation of several components of the transposon-based vector. The present invention further includes vectors containing more than one gene of interest, wherein a second or subsequent gene of interest is operably-linked to the second promoter or to a different promoter. It is also to be understood that the transposon-based vectors shown in the Figures are representative of the present invention and that the order of the vector elements may be different than that shown in the Figures, that the elements may be present in various orientations, and that the vectors may contain additional elements not shown in the Figures.

### Transposases and Insertion Sequences

In a further embodiment of the present invention, the transposase found in the transposase-based vector is an altered target site (ATS) transposase and the insertion sequences are those recognized by the ATS transposase. However, the transposase located in the transposase-based vectors is not limited to a modified ATS transposase and can be derived from any transposase. Transposases known in the prior art include those found in AC7, Tn5SEQ1, Tn916, Tn951, Tn1721, Tn 2410, Tn1681, Tn1, Tn2, Tn3, Tn4, Tn5, Tn6, Tn9, Tn10, Tn30, Tn101, Tn903, Tn501, Tn1000 ( $\gamma\delta$ ), Tn1681, Tn2901, AC transposons, Mp transposons, Spm transposons, En transposons, Dotted transposons, Mu transposons, Ds transposons, dSpm transposons and I transposons. According to the present invention, these transposases and their regulatory sequences are modified for improved functioning as follows: a) the addition one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:8) at the 3' end of the promoter operably-linked to the transposase; b) a change of the codons for the first several amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) the addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) the addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene.

Although not wanting to be bound by the following statement, it is believed that the modifications of the first several N-terminal codons of the transposase gene increase transcription of the transposase gene, in part, by increasing strand dissociation. It is preferable that between approximately 1 and 20, more preferably 3 and 15, and most preferably between 4 and 12 of the first N-terminal codons of the transposase are modified such that the third base of each codon is changed to an A or a T without changing the encoded amino acid. In one embodiment, the first ten N-terminal codons of the transposase gene are modified in this manner. It is also preferred that the transposase contain mutations that make it less specific for preferred insertion sites and thus increases the rate of transgene insertion as discussed in U.S. Patent No. 5,719,055.

In some embodiments, the transposon-based vectors are optimized for expression in a particular host by changing the methylation patterns of the vector

DNA. For example, prokaryotic methylation may be reduced by using a methylation deficient organism for production of the transposon-based vector. The transposon-based vectors may also be methylated to resemble eukaryotic DNA for expression in a eukaryotic host.

5 Transposases and insertion sequences from other analogous eukaryotic transposon-based vectors that can also be modified and used are, for example, the *Drosophila* P element derived vectors disclosed in U.S. Patent No. 6,291,243; the *Drosophila* mariner element described in Sherman et al. (1998); or the sleeping beauty transposon. See also Hackett et al. (1999); D. Lampe et al., 1999. Proc. Natl. Acad. Sci. USA, 96:11428-11433; S. Fischer et al., 2001. Proc. Natl. Acad. Sci. USA, 98:6759-6764; L. Zagoraïou et al., 2001. Proc. Natl. Acad. Sci. USA, 98:11474-11478; and D. Berg et al. (Eds.), *Mobile DNA*, Amer. Soc. Microbiol. (Washington, D.C., 1989). However, it should be noted that bacterial transposon-based elements are preferred, as there is less likelihood that a eukaryotic transposase in the recipient species will recognize prokaryotic insertion sequences bracketing the transgene.

10 Many transposases recognize different insertion sequences, and therefore, it is to be understood that a transposase-based vector will contain insertion sequences recognized by the particular transposase also found in the transposase-based vector. In a preferred embodiment of the invention, the insertion sequences have been shortened to about 70 base pairs in length as compared to those found in wild-type transposons that typically contain insertion sequences of well over 100 base pairs.

While the examples provided below incorporate a "cut and insert" Tn10 based vector that is destroyed following the insertion event, the present invention also encompasses the use of a "rolling replication" type transposon-based vector. Use of a rolling replication type transposon allows multiple copies of the transposon/transgene to be made from a single transgene construct and the copies inserted. This type of transposon-based system thereby provides for insertion of multiple copies of a transgene into a single genome. A rolling replication type transposon-based vector may be preferred when the promoter operably-linked to gene of interest is endogenous to the host cell and present in a high copy number or highly expressed. However, use of a rolling replication system may require tight control to limit the insertion events to non-lethal levels. Tn1, Tn2, Tn3, Tn4, Tn5, Tn9, Tn21, Tn501, Tn551, Tn951, Tn1721, Tn2410 and Tn2603 are examples of a rolling replication type transposon, although Tn5 could be both a rolling replication and a cut and insert type transposon.

### Stop Codons and PolyA Sequences

In one embodiment, the transposon-based vector contains two stop codons operably-linked to the transposase and/or to the gene of interest. In an alternate  
5 embodiment, one stop codon of UAA or UGA is operably linked to the transposase and/or to the gene of interest. While not wanting to be bound by the following statement, it is thought that the stop codon UAG is less effective in translation termination and is therefore less desirable in the constructs described herein.

As used herein an "effective polyA sequence" refers to either a synthetic or  
10 non-synthetic sequence that contains multiple and sequential nucleotides containing an adenine base (an A polynucleotide string) and that increases expression of the gene to which it is operably-linked. A polyA sequence may be operably-linked to any gene in the transposon-based vector including, but not limited to, a transposase gene and a gene of interest. A preferred polyA sequence is optimized for use in the host animal  
15 or human and for the desired end product.

The goal is to use a poly A that gives a similar level of expression as the gene being replaced, or the desired result. With siRNA for example, only a few copies of the RNAi sequence are required, so the mRNA may not have to be extremely stable, and in fact may be detrimental or just a waste of energy for the cell (See Zhang et al.,  
20 Nucleic Acids Research, database issue, Vol. 33:D116-D120 2005).

In one embodiment, the polyA sequence is optimized for use in an avian species and more specifically, a chicken. An avian optimized polyA sequence generally contains a minimum of 40 base pairs, preferably between approximately 40 and several hundred base pairs, and more preferably approximately 75 base pairs that  
25 precede the A polynucleotide string and thereby separate the stop codon from the A polynucleotide string. In one embodiment of the present invention, the polyA sequence comprises a conalbumin polyA sequence as provided in SEQ ID NO:17 and as taken from GenBank accession # Y00407, base pairs 10651-11058. In another embodiment, the polyA sequence comprises a synthetic polynucleotide sequence  
30 shown in SEQ ID NO:18. In yet another embodiment, the polyA sequence comprises an avian optimized polyA sequence provided in SEQ ID NO:19. A chicken optimized polyA sequence may also have a reduced amount of CT repeats as compared to a synthetic polyA sequence.

It is a surprising discovery of the present invention that such an avian optimized poly A sequence increases expression of a polynucleotide to which it is operably-linked in an avian as compared to a non-avian optimized polyA sequence. It is to be understood that polyA sequences may be optimized for other classes of animals, such as mammals, and used in the transposon-based vectors of the present invention to provide gene therapy. Accordingly, the present invention includes methods of or increasing incorporation of a gene of interest wherein the gene of interest resides in a transposon-based vector containing a transposase gene and wherein the transposase gene is operably linked to an avian optimized polyA sequence. The present invention also includes methods of increasing expression of a gene of interest in an avian that includes administering a gene of interest to the avian, wherein the gene of interest is operably-linked to an avian optimized polyA sequence. An avian optimized polyA nucleotide string is defined herein as a polynucleotide containing an A polynucleotide string and a minimum of 40 base pairs, preferably between approximately 40 and several hundred base pairs, and more preferably approximately 75 base pairs that precede the A polynucleotide string. The present invention further provides transposon-based vectors containing a gene of interest or transposase gene operably linked to an avian optimized polyA sequence.

## Promoters and Enhancers

The first promoter operably-linked to the transposase gene and the second promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. Constitutive promoters include, but are not limited to, immediate early cytomegalovirus (CMV) promoter, herpes simplex virus 1 (HSV1) immediate early promoter, SV40 promoter, lysozyme promoter, early and late CMV promoters, early and late HSV promoters,  $\beta$ -actin promoter, tubulin promoter, Rous-Sarcoma virus (RSV) promoter, and heat-shock protein (HSP) promoter. Inducible promoters include tissue-specific promoters, developmentally-regulated promoters, and chemically inducible promoters. Examples of tissue-specific promoters include the glucose-6-phosphatase (G6P) promoter, vitellogenin promoter, ovalbumin promoter, ovomucoid promoter, conalbumin promoter, ovotransferrin promoter, prolactin promoter, kidney uromodulin promoter, and placental lactogen promoter. In one embodiment, the vitellogenin promoter includes a polynucleotide sequence of SEQ ID

NO:20. The G6P promoter sequence may be deduced from a rat G6P gene untranslated upstream region provided in GenBank accession number U57552.1. Examples of developmentally-regulated promoters include the homeobox promoters and several hormone induced promoters. Examples of chemically inducible  
5 promoters include reproductive hormone induced promoters and antibiotic inducible promoters such as the tetracycline inducible promoter and the zinc-inducible metallothionine promoter.

Other inducible promoter systems include the Lac operator repressor system inducible by IPTG (isopropyl beta-D-thiogalactoside) (Cronin, A. et al. 2001. Genes  
10 and Development, v. 15), ecdysone-based inducible systems (Hoppe, U. C. et al. 2000. Mol. Ther. 1:159-164); estrogen-based inducible systems (Braselmann, S. et al. 1993. Proc. Natl. Acad. Sci. 90:1657-1661); progesterone-based inducible systems using a chimeric regulator, GLVP, which is a hybrid protein consisting of the GAL4 binding domain and the herpes simplex virus transcriptional activation domain, VP16,  
15 and a truncated form of the human progesterone receptor that retains the ability to bind ligand and can be turned on by RU486 (Wang, et al. 1994. Proc. Natl. Acad. Sci. 91:8180-8184); CID-based inducible systems using chemical inducers of dimerization (CIDs) to regulate gene expression, such as a system wherein rapamycin induces dimerization of the cellular proteins FKBP12 and FRAP (Belshaw, P. J. et al. 1996. J.  
20 Chem. Biol. 3:731-738; Fan, L. et al. 1999. Hum. Gene Ther. 10:2273-2285; Shariat, S.F. et al. 2001. Cancer Res. 61:2562-2571; Spencer, D.M. 1996. Curr. Biol. 6:839-847). Chemical substances that activate the chemically inducible promoters can be administered to the animal containing the transgene of interest via any method known to those of skill in the art.

25 Other examples of cell-specific and constitutive promoters include but are not limited to smooth-muscle SM22 promoter, including chimeric SM22alpha/telokin promoters (Hoggatt A.M. et al., 2002. Circ Res. 91(12):1151-9); ubiquitin C promoter (Biochim Biophys Acta, 2003. Jan. 3:1625(1):52-63); Hsf2 promoter; murine COMP (cartilage oligomeric matrix protein) promoter; early B cell-specific mb-1 promoter  
30 (Sigvardsson M., et al., 2002. Mol. Cell Biol. 22(24):8539-51); prostate specific antigen (PSA) promoter (Yoshimura I. et al., 2002, J. Urol. 168(6):2659-64); exorh promoter and pineal expression-promoting element (Asaoka Y., et al., 2002. Proc. Natl. Acad. Sci. 99(24):15456-61); neural and liver ceramidase gene promoters (Okino N. et al., 2002. Biochem. Biophys. Res. Commun. 299(1):160-6); PSP94 gene



promoter/enhancer (Gabril M.Y. et al., 2002. *Gene Ther.* 9(23):1589-99); promoter of the human FAT/CD36 gene (Kuriki C., et al., 2002. *Biol. Pharm. Bull.* 25(11):1476-8); VL30 promoter (Staplin W.R. et al., 2002. *Blood* October 24, 2002); and, IL-10 promoter (Brenner S., et al., 2002. *J. Biol. Chem.* December 18, 2002). Additional promoters are shown in Table 1.

Examples of avian promoters include, but are not limited to, promoters controlling expression of egg white proteins, such as ovalbumin, ovotransferrin (conalbumin), ovomucoid, lysozyme, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, ovostatin (ovomacroglobin), cystatin, avidin, thiamine-binding protein, glutamyl aminopeptidase minor glycoprotein 1, minor glycoprotein 2; and promoters controlling expression of egg-yolk proteins, such as vitellogenin, very low-density lipoproteins, low density lipoprotein, cobalamin-binding protein, riboflavin-binding protein, biotin-binding protein (Awade, 1996. *Z. Lebensm. Unters. Forsch.* 202:1-14). An advantage of using the vitellogenin promoter is that it is active during the egg-laying stage of an animal's life-cycle, which allows for the production of the protein of interest to be temporally connected to the import of the protein of interest into the egg yolk when the protein of interest is equipped with an appropriate targeting sequence. In some embodiments, the avian promoter is an oviduct-specific promoter. As used herein, the term "oviduct-specific promoter" includes, but is not limited to, ovalbumin; ovotransferrin (conalbumin); ovomucoid; 01, 02, 03, 04 or 05 avidin; ovomucin; g2 ovoglobulin; g3 ovoglobulin; ovoflavoprotein; and ovostatin (ovomacroglobin) promoters.

When germline transformation occurs via intraovarian or intratesticular administration, or when hepatocytes are targeted for incorporation of components of a vector through non-germ line administration, liver-specific promoters may be operably-linked to the gene of interest to achieve liver-specific expression of the transgene. Liver-specific promoters of the present invention include, but are not limited to, the following promoters, vitellogenin promoter, G6P promoter, cholesterol-7-alpha-hydroxylase (CYP7A) promoter, phenylalanine hydroxylase (PAH) promoter, protein C gene promoter, insulin-like growth factor I (IGF-I) promoter, bilirubin UDP-glucuronosyltransferase promoter, aldolase B promoter, furin promoter, metallothionein promoter, albumin promoter, and insulin promoter.

Also included in the present invention are promoters that can be used to target expression of a protein of interest into the milk of a milk-producing animal including,

but not limited to,  $\beta$  lactoglobulin promoter, whey acidic protein promoter, lactalbumin promoter and casein promoter.

When germline transformation occurs via intraovarian or intratesticular administration, or when cells of the immune system are targeted through non-germ line administration, immune system-specific promoters may be operably-linked to the gene of interest to achieve immune system-specific expression of the transgene. Accordingly, promoters associated with cells of the immune system may also be used. Acute phase promoters such as interleukin (IL)-1 and IL-2 may be employed. Promoters for heavy and light chain Ig may also be employed. The promoters of the T cell receptor components CD4 and CD8, B cell promoters and the promoters of CR2 (complement receptor type 2) may also be employed. Immune system promoters are preferably used when the desired protein is an antibody protein.

It is to be understood that any cell may be targeted for incorporation of a desired gene to provide gene therapy. Such cells may include, without limitation, endocrine cells or cancer cells. Promoters specific for selected endocrine cells, such as insulin-producing islet cells, hypophyseal growth hormone producing cells, or estrogen-producing follicular cells are known to one of ordinary skill in the art and may be incorporated into the transposon-based vectors of the present invention in order to provide gene therapy, by modulating hormone synthesis and secretion. Endocrine disorders and associated conditions of over or underproduction of hormones are known to one of ordinary skill in the art and many such conditions are described in textbooks such as Williams Textbook of Endocrinology, 10<sup>th</sup> ed., Williams, R.H. et al., eds. 2002, W.B. Saunders. Specific cancerous cells, such as ovarian cancer cells, are known to produce and release specific molecules. Promoters specific for these cells, and other cancerous cells, are known to one of ordinary skill in the art and may be incorporated into the transposon-based vectors of the present invention in order to provide gene therapy, perhaps by producing inhibitory RNA in these cells or by producing proteins or peptides that interfere with cancer cell function or replication.

Additional gene targets, especially related to cancer, are oncogenes and also genes involved in cellular functions such as cell division, microtubule production and spindle formation, growth factors, growth factor receptors, oncoproteins and signal transduction pathways associated with transducing signals associated with function of cancerous cells. It is to be understood that a gene target may be a particular protein or

enzyme involved in a multistep process associate with cell function. For example, an enzyme that is a component in a metabolic pathway of several enzymes may be disrupted using inhibitory RNA, thereby affecting the function of this metabolic pathway.

5        Also included in this invention are modified promoters/enhancers wherein elements of a single promoter are duplicated, modified, or otherwise changed. In one embodiment, steroid hormone-binding domains of the ovalbumin promoter are moved from about -3.5 kb to within approximately the first 1000 base pairs of the gene of interest. Modifying an existing promoter with promoter/enhancer elements not found  
10 naturally in the promoter, as well as building an entirely synthetic promoter, or drawing promoter/enhancer elements from various genes together on a non-natural backbone, are all encompassed by the current invention.

Accordingly, it is to be understood that the promoters contained within the transposon-based vectors of the present invention may be entire promoter sequences  
15 or fragments of promoter sequences. For example, in one embodiment, the promoter operably linked to a gene of interest is an approximately 900 base pair fragment of a chicken ovalbumin promoter (SEQ ID NO:21). The constitutive and inducible promoters contained within the transposon-based vectors may also be modified by the addition of one or more modified Kozak sequences of ACCATG (SEQ ID NO:8).

20        As indicated above, the present invention includes transposon-based vectors containing one or more enhancers. These enhancers may or may not be operably-linked to their native promoter and may be located at any distance from their operably-linked promoter. A promoter operably-linked to an enhancer and a promoter modified to eliminate repressive regulatory effects are referred to herein as an  
25 “enhanced promoter.” The enhancers contained within the transposon-based vectors may be enhancers found in birds, such as an ovalbumin enhancer, but are not limited to these types of enhancers. In one embodiment, an approximately 675 base pair enhancer element of an ovalbumin promoter is cloned upstream of an ovalbumin promoter with 300 base pairs of spacer DNA separating the enhancer and promoter.  
30 In one embodiment, the enhancer used as a part of the present invention comprises base pairs 1-675 of a chicken ovalbumin enhancer from GenBank accession #S82527.1. The polynucleotide sequence of this enhancer is provided in SEQ ID NO:22.

Also included in some of the transposon-based vectors of the present invention are cap sites and fragments of cap sites. In one embodiment, approximately 50 base pairs of a 5' untranslated region wherein the capsite resides are added on the 3' end of an enhanced promoter or promoter. An exemplary 5' untranslated region is provided  
5 in SEQ ID NO:23. A putative cap-site residing in this 5' untranslated region preferably comprises the polynucleotide sequence provided in SEQ ID NO:24.

In one embodiment of the present invention, the first promoter operably-linked to the transposase gene is a constitutive promoter and the second promoter operably-linked to the gene of interest is a cell specific promoter. In the second embodiment,  
10 use of the first constitutive promoter allows for constitutive activation of the transposase gene and incorporation of the gene of interest into virtually all cell types, including the germline of the recipient animal. Although the gene of interest is incorporated into the germline generally, the gene of interest may only be expressed in a tissue-specific manner to achieve gene therapy. A transposon-based vector  
15 having a constitutive promoter operably-linked to the transposase gene can be administered by any route, and in one embodiment, the vector is administered to an ovary, to an artery leading to the ovary or to a lymphatic system or fluid proximal to the ovary. In another embodiment, the transposon-based vector having a constitutive promoter operably-linked to the transposase gene can be administered to vessels  
20 supplying the liver, muscle, brain, lung, kidney, heart or any other desired organ, tissue or cellular target.

It should be noted that cell- or tissue-specific expression as described herein does not require a complete absence of expression in cells or tissues other than the preferred cell or tissue. Instead, "cell-specific" or "tissue-specific" expression refers  
25 to a majority of the expression of a particular gene of interest in the preferred cell or tissue, respectively.

When incorporation of the gene of interest into the germline is not preferred, the first promoter operably-linked to the transposase gene can be a tissue-specific promoter. For example, transfection of a transposon-based vector containing a  
30 transposase gene operably-linked to a liver specific promoter such as the G6P promoter or vitellogenin promoter provides for activation of the transposase gene and incorporation of the gene of interest in the cells of the liver but not into the germline and other cells generally. In another example, transfection of a transposon-based vector containing a transposase gene operably-linked to an oviduct specific promoter

such as the ovalbumin promoter provides for activation of the transposase gene and incorporation of the gene of interest in the cells of the oviduct but not into the germline and other cells generally. In this embodiment, the second promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. In one embodiment, both the first promoter and the second promoter are an ovalbumin promoter. In embodiments wherein tissue-specific expression or incorporation is desired, it is preferred that the transposon-based vector is administered directly to the tissue of interest, to an artery leading to the organ or tissue of interest or to fluids surrounding the organ or tissue of interest. In one embodiment, the tissue of interest is the oviduct and administration is achieved by direct injection into the oviduct or an artery leading to the oviduct. In another embodiment, the tissue of interest is the liver and administration is achieved by direct injection into the portal vein or hepatic artery. In another embodiment, the tissue of interest is cardiac muscle tissue in the heart and administration is achieved by direct injection into the coronary arteries. In another embodiment, the tissue of interest is neural tissue and administration is achieved by direct injection into a cerebrovascular or spinovascular artery. In yet another embodiment, the target is a solid tumor and the administration is achieved by injection into a vessel supplying the tumor or by injection into the tumor. In yet another embodiment, the target is a diffuse cancer such as ovarian cancer spread throughout the abdominopelvic cavity and the administration is achieved by injection into the abdominopelvic cavity. In yet another embodiment, the target is the lung, for example the surfactant-producing cells, and the administration is achieved by injection into the right ventricle, the pulmonary artery or a branch thereof, or by aerosol administration into the respiratory system. In still another embodiment, the target is a lymph node and the administration is achieved by injection into lymphatic vessels supplying that node.

Accordingly, cell specific promoters may be used to enhance transcription in selected tissues. In birds, for example, promoters that are found in cells of the fallopian tube, such as ovalbumin, conalbumin, ovomucoid and/or lysozyme, are used in the vectors to ensure transcription of the gene of interest in the epithelial cells and tubular gland cells of the fallopian tube, leading to synthesis of the desired protein encoded by the gene and deposition into the egg white. In mammals, promoters specific for the epithelial cells of the alveoli of the mammary gland, such as prolactin, insulin, beta lactoglobulin, whey acidic protein, lactalbumin, casein, and/or placental

lactogen, are used in the design of vectors used for transfection of these cells for the production of desired proteins for deposition into the milk. In liver cells, the G6P promoter may be employed to drive transcription of the gene of interest for protein production. Proteins made in the liver of birds may be delivered to the egg yolk.

- 5           In order to achieve higher or more efficient expression of the transposase gene, the promoter and other regulatory sequences operably-linked to the transposase gene may be those derived from the host. These host specific regulatory sequences can be tissue specific as described above or can be of a constitutive nature. For example, an avian actin promoter and its associated polyA sequence can be operably-
- 10       linked to a transposase in a transposase-based vector for transfection into an avian. Examples of other host specific promoters that could be operably-linked to the transposase include the myosin and DNA or RNA polymerase promoters.

#### Directing Sequences

- 15           In some embodiments of the present invention, the gene of interest is operably-linked to a directing sequence or a sequence that provides proper conformation to the desired protein encoded by the gene of interest. As used herein, the term "directing sequence" refers to both signal sequences and targeting sequences. An egg directing sequence includes, but is not limited to, an ovomucoid signal
- 20       sequence, an ovalbumin signal sequence, a cecropin pre pro signal sequence, and a vitellogenin targeting sequence. The term "signal sequence" refers to an amino acid sequence, or the polynucleotide sequence that encodes the amino acid sequence, that directs the protein to which it is linked to the endoplasmic reticulum in a eukaryote, and more preferably the translocational pores in the endoplasmic reticulum, or the
- 25       plasma membrane in a prokaryote, or mitochondria, such as for the purpose of gene therapy for mitochondrial diseases. Signal and targeting sequences can be used to direct a desired protein into, for example, the bloodstream, when the transposon-based vectors are administered to the liver of an animal.

- Signal sequences can also be used to direct a desired protein into, for example,
- 30       a secretory pathway for secretion and release of the desired protein. For example appropriate signal sequences can be employed to provide gene therapy for enhanced secretion of growth hormone from selected cells, for example growth hormone cells, or liver cells. This therapy is useful for treating deficiencies in circulating growth

hormone levels and reduced stature. Liver specific promoters may be used to enhance production and release of antibodies or hepatic proteins such as globulins.

Signal sequences can also be used to direct a desired protein into, for example, a secretory pathway for incorporation into the egg yolk or the egg white, when the transposon-based vectors are administered to a bird or other egg-laying animal. One example of such a transposon-based vector is provided in Figure 3 wherein the gene of interest is operably linked to the ovomucoid signal sequence. The present invention also includes a gene of interest operably-linked to a second gene containing a signal sequence. An example of such an embodiment is shown in Figure 2 wherein the gene of interest is operably-linked to the ovalbumin gene that contains an ovalbumin signal sequence. Other signal sequences that can be included in the transposon-based vectors include, but are not limited to the ovotransferrin and lysozyme signal sequences. In one embodiment, the signal sequence is an ovalbumin signal sequence including a sequence shown in SEQ ID NO:25. In another embodiment, the signal sequence is a modified ovalbumin signal sequence including a sequence shown in SEQ ID NO:26 or SEQ ID NO:27.

As also used herein, the term "targeting sequence" refers to an amino acid sequence, or the polynucleotide sequence encoding the amino acid sequence, which amino acid sequence is recognized by a receptor located on the exterior of a cell. Binding of the receptor to the targeting sequence results in uptake of the protein or peptide operably-linked to the targeting sequence by the cell. One example of a targeting sequence is a vitellogenin targeting sequence that is recognized by a vitellogenin receptor (or the low density lipoprotein receptor) on the exterior of an oocyte. In one embodiment, the vitellogenin targeting sequence includes the polynucleotide sequence of SEQ ID NO:28. In another embodiment, the vitellogenin targeting sequence includes all or part of the vitellogenin gene. Other targeting sequences include VLDL and Apo E, which are also capable of binding the vitellogenin receptor. Since the ApoE protein is not endogenously expressed in birds, its presence may be used advantageously to identify birds carrying the transposon-based vectors of the present invention.

#### Genes of Interest Encoding Desired Proteins

A gene of interest selected for stable incorporation is designed to encode any desired protein or peptide or nucleic acid or to regulate any cellular response. In some

embodiments, the desired proteins or peptides are released from cells into the surrounding environment, the lymphatic system or the vascular system. In some embodiments, the desired proteins or peptides are deposited in an egg or in milk. In other embodiments the desired proteins or peptides may be directed to an axon, to a  
5 hepatocyte cell membrane for release into the bloodstream, to the membrane of a beta lymphocyte for release into the circulation. It is to be understood that the present invention encompasses transposon-based vectors containing multiple genes of interest. The multiple genes of interest may each be operably-linked to a separate promoter and other regulatory sequence(s) or may all be operably-linked to the same  
10 promoter and other regulatory sequences(s). In one embodiment, multiple gene of interest are linked to a single promoter and other regulatory sequence(s) and each gene of interest is separated by a cleavage site or a portion of a signal sequence. A gene of interest may contain modifications of the codons for the first several N-terminal amino acids of the gene of interest, wherein the third base of each codon is  
15 changed to an A or a T without changing the corresponding amino acid.

Protein and peptide hormones are a preferred class of proteins in the present invention. Such protein and peptide hormones are synthesized throughout the endocrine system and include, but are not limited to, hypothalamic hormones and hypophysiotropic hormones, anterior, intermediate and posterior pituitary hormones,  
20 pancreatic islet hormones, hormones made in the gastrointestinal system, renal hormones, thymic hormones, parathyroid hormones, adrenal cortical and medullary hormones. Specifically, hormones that can be produced using the present invention include, but are not limited to, chorionic gonadotropin, corticotropin, erythropoietin, glucagons, IGF-1, oxytocin, platelet-derived growth factor, calcitonin, follicle-stimulating hormone, luteinizing hormone, thyroid-stimulating hormone, insulin,  
25 gonadotropin-releasing hormone and its analogs, vasopressin, octreotide, somatostatin, prolactin, adrenocorticotrophic hormone, antidiuretic hormone, thyrotropin-releasing hormone (TRH), growth hormone-releasing hormone (GHRH), parathyroid hormone (PTH), glucagons, calcitriol, calciferol, atrial-natriuretic peptide, gastrin, secretin, cholecystokinin (CCK), neuropeptide Y, ghrelin, PYY<sub>3-36</sub>,  
30 angiotensinogen, thrombopoietin, and leptin. It is to be understood that proteins that are normally folded or have chains or component parts that combine to form the active protein may be produced in a linear manner, for example luteinizing hormone (LH).



Other multimeric proteins that may be produced using the present invention are as follows: factors involved in the synthesis or replication of DNA, such as DNA polymerase alpha and DNA polymerase delta; proteins involved in the production of mRNA, such as TFIID and TFIIF; cell, nuclear and other membrane-associated proteins, such as hormone and other signal transduction receptors, active transport proteins and ion channels, multimeric proteins in the blood, including hemoglobin, fibrinogen and von Willibrand's Factor; proteins that form structures within the cell, such as actin, myosin, and tubulin and other cytoskeletal proteins; proteins that form structures in the extra cellular environment, such as collagen, elastin and fibronectin; proteins involved in intra- and extra-cellular transport, such as kinesin and dynein, the SNARE family of proteins (soluble NSF attachment protein receptor) and clathrin; proteins that help regulate chromatin structure, such as histones and protamines, Swi3p, Rsc8p and moira; multimeric transcription factors such as Fos, Jun and CBTF (CCAAT box transcription factor); multimeric enzymes such as acetylcholinesterase and alcohol dehydrogenase; chaperone proteins such as GroE, Gro EL (chaperonin 60) and Gro ES (chaperonin 10); anti-toxins, such as snake venom, botulism toxin, Streptococcus super antigens; lysins (enzymes from bacteriophage and viruses); as well as most allosteric proteins. By using appropriate polynucleotide sequences, species-specific hormones may be made by transgenic animals.

In one embodiment of the present invention, the gene of interest is a proinsulin gene and the desired molecule is insulin. Proinsulin consists of three parts: a C-peptide and two strands of amino acids (the alpha and beta chains) that later become linked together to form the insulin molecule. Figures 2 and 3 are schematics of transposon-based vector constructs containing a proinsulin gene operably-linked to an ovalbumin promoter and ovalbumin protein or an ovomucoid promoter and ovomucoid signal sequence, respectively. In these embodiments, proinsulin is expressed in the oviduct tubular gland cells and then deposited in the egg white. One example of a proinsulin polynucleotide sequence is shown in SEQ ID NO:29, wherein the C-peptide cleavage site spans from Arg at position 31 to Arg at position 65. In other embodiments, the construct is designed for stable incorporation into hepatocytes and production of insulin for release into the vascular system.

In another embodiment of the present invention a vector is constructed for use in gene therapy for diabetes (Figure 6). A hepatocyte specific promoter is placed upstream of the transposase (ATS). The hepatocyte specific promoter could be a

glucose-6-phosphatase promoter, liver specific albumin promoter, serum alpha-fetoprotein promoter, or other hepatocyte specific promoter. Such a specific promoter permits gene incorporation into the genome of hepatocytes since the transposase would not be expressed in other cell types. The promoter driving expression of the proinsulin gene is more specific, such as the glucose-6-phosphatase promoter. This promoter is desirable because it responds to blood glucose levels similar to beta islet cells of the pancreas. The proinsulin also contains a signal sequence to allow secretion from the liver cell. For instance, a signal sequence can be an albumin signal sequence or alpha-fetoprotein signal sequence. Likewise, the poly A is from a liver specific protein in order to optimize mRNA stability for the amount of desired expression. This is easily determined by one skilled in the art. The proinsulin and liver specific sequences are from the species of animal targeted for gene therapy, i.e., human sequence for human gene therapy, canine sequence for canine gene therapy or feline sequences for gene therapy in felines.

In another embodiment of the present invention a vector is constructed for use in gene therapy for treatment of growth hormone deficiency by expressing growth hormone from hepatocytes (Figure 7). A liver specific promoter limits incorporation of the gene to hepatocytes. To further limit expression to hepatocytes, the vector is delivered as linear DNA as opposed to supercoiled DNA. Linear DNA has the added advantage of being destroyed more quickly than supercoiled DNA, so that if the DNA were delivered to a cell and the promoter was leaky (a low basal level of expression), the chances of expression before degradation would be minimized. The selection of growth hormone expression level is related to the dosage desired, i.e. strong constitutive promoter for larger doses, low to intermediate constitutive promoter for smaller doses. The signal sequence and poly A are hepatocyte derived for proper secretion and mRNA stability, respectively.

In another embodiment of the present invention a vector is constructed for use in gene therapy for treatment of cystic fibrosis by specifically expressing the normal CFTR gene in respiratory epithelial cells (Figure 8). To incorporate the desired transgene into respiratory epithelial cells, the ciliated cell-specific promoter (FOXJ1), or another lung specific promoter, is used to drive expression of the transposase. To treat the disease, a normal CFTR gene is delivered to the respiratory epithelial cells.

In another embodiment of the present invention a vector is constructed for use in gene therapy for treatment of cancer by specifically expressing the cholera toxin

gene in cancer cells (Figure 9). By linking the transposase to a cancer specific promoter, only cancer cells are stably transformed with the target gene. The target gene can encode for a toxin, such as cholera toxin A, expressed constitutively or under control of a cancer specific promoter to selectively kill the transfected cancer cells. In another embodiment, the transposase is placed under control of a cell specific promoter so that only one cell is transformed. In one embodiment, the target gene is a secreted fusion peptide, such as a peptide that has a component recognized by a surface receptor on a cancer cell and a lytic component that would destroy the cell following the binding of the other part of the fusion peptide to the cell surface receptor. In one embodiment, the target gene could encode for betaLH/Phor14, which is a ligand/lytic peptide combination that targets a receptor on a cancer cell with LH receptors and kills that cell with little or no damage to surrounding healthy tissue.

Serum proteins including lipoproteins such as high density lipoprotein (HDL), HDL-Milano and low density lipoprotein, apolipoprotein, albumin, clotting cascade factors, factor VIII, factor IX, fibrinogen, and globulins are also included in the group of desired proteins of the present invention. Immunoglobulins are one class of desired globulin molecules and include but are not limited to IgG, IgM, IgA, IgD, IgE, IgY, lambda chains, kappa chains and fragments thereof; bi-specific antibodies, and fragments thereof; scFv fragments, Fc fragments, and Fab fragments as well as dimeric, trimeric and oligomeric forms of antibody fragments. Desired antibodies include, but are not limited to, naturally occurring antibodies, animal-specific antibodies, human antibodies, humanized antibodies, autoantibodies and hybrid antibodies. Genes encoding modified versions of naturally occurring antibodies or fragments thereof and genes encoding artificially designed antibodies or fragments thereof may be incorporated into the transposon-based vectors of the present invention. Desired antibodies also include antibodies with the ability to bind specific ligands, for example, antibodies against proteins associated with cancer-related molecules, such as anti-her 2, or anti-CA125. Accordingly, the present invention encompasses a transposon-based vector containing one or more genes encoding a heavy immunoglobulin (Ig) chain and a light Ig chain. Further, more than one gene encoding for more than one antibody may be administered in one or more transposon-based vectors of the present invention. In this manner, antibodies may be made in liver cells or another cell selected for transfection, such as fibroblasts and released locally or gain access to the circulation. In one embodiment, a transposon-based

vector contains a heavy Ig chain and a light Ig chain, both operably linked to a promoter.

Antibodies used as therapeutic reagents include but are not limited to antibodies for use in cancer immunotherapy against specific antigens, or for providing  
5 passive immunity to an animal against an infectious disease or a toxic agent. Antibodies may be made by the animal receiving the transposon-based vectors to facilitate the animal's immune response to a selected antigen. Animals receiving gene therapy to enhance resistance to a disease or to fight an ongoing disease, such as cancer, may receive a transposon-based vector containing genes encoding antibodies  
10 that bind to epitopes on cancer cells.

Antibodies that may be made with the practice of the present invention include, but are not limited to primary antibodies, secondary antibodies, designer antibodies, anti-protein antibodies, anti-peptide antibodies, anti-DNA antibodies, anti-RNA antibodies, anti-hormone antibodies, anti-hypophysiotropic peptides, antibodies  
15 against non-natural antigens, anti-anterior pituitary hormone antibodies, anti-posterior pituitary hormone antibodies, anti-venom antibodies, anti-tumor marker antibodies, antibodies directed against epitopes associated with infectious disease, including, anti-viral, anti-bacterial, anti-protozoal, anti-fungal, anti-parasitic, anti-receptor, anti-lipid, anti-phospholipid, anti-growth factor, anti-cytokine, anti-monokine, anti-idiotypic, and  
20 anti-accessory (presentation) protein antibodies. Antibodies made with the present invention, as well as light chains or heavy chains, may also be used to inhibit enzyme activity.

Antibodies that may be produced using the present invention include, but are not limited to, antibodies made against the following proteins: Bovine  $\gamma$ -Globulin,  
25 Serum; Bovine IgG, Plasma; Chicken  $\gamma$ -Globulin, Serum; Human  $\gamma$ -Globulin, Serum; Human IgA, Plasma; Human IgA<sub>1</sub>, Myeloma; Human IgA<sub>2</sub>, Myeloma; Human IgA<sub>2</sub>, Plasma; Human IgD, Plasma; Human IgE, Myeloma; Human IgG, Plasma; Human IgG, Fab Fragment, Plasma; Human IgG, F(ab')<sub>2</sub> Fragment, Plasma; Human IgG, Fc Fragment, Plasma; Human IgG<sub>1</sub>, Myeloma; Human IgG<sub>2</sub>, Myeloma; Human IgG<sub>3</sub>,  
30 Myeloma; Human IgG<sub>4</sub>, Myeloma; Human IgM, Myeloma; Human IgM, Plasma; Human Immunoglobulin, Light Chain  $\kappa$ , Urine; Human Immunoglobulin, Light Chains  $\kappa$  and  $\lambda$ , Plasma; Mouse  $\gamma$ -Globulin, Serum; Mouse IgG, Serum; Mouse IgM, Myeloma; Rabbit  $\gamma$ -Globulin, Serum; Rabbit IgG, Plasma; and Rat  $\gamma$ -Globulin, Serum. In one embodiment, the transposon-based vector comprises the coding

sequence of light and heavy chains of a murine monoclonal antibody that shows specificity for human seminoprotein (GenBank Accession numbers AY129006 and AY129304 for the light and heavy chains, respectively).

A further non-limiting list of antibodies that recognize other antibodies is as follows: Anti-Chicken IgG, heavy (H) & light (L) Chain Specific (Sheep); Anti-Goat  $\gamma$ -Globulin (Donkey); Anti-Goat IgG, Fc Fragment Specific (Rabbit); Anti-Guinea Pig  $\gamma$ -Globulin (Goat); Anti-Human Ig, Light Chain, Type  $\kappa$  Specific; Anti-Human Ig, Light Chain, Type  $\lambda$  Specific; Anti-Human IgA,  $\alpha$ -Chain Specific (Goat); Anti-Human IgA, Fab Fragment Specific; Anti-Human IgA, Fc Fragment Specific; Anti-Human IgA, Secretory; Anti-Human IgE,  $\epsilon$ -Chain Specific (Goat); Anti-Human IgE, Fc Fragment Specific; Anti-Human IgG, Fc Fragment Specific (Goat); Anti-Human IgG,  $\gamma$ -Chain Specific (Goat); Anti-Human IgG, Fc Fragment Specific; Anti-Human IgG, Fd Fragment Specific; Anti-Human IgG, H & L Chain Specific (Goat); Anti-Human IgG<sub>1</sub>, Fc Fragment Specific; Anti-Human IgG<sub>2</sub>, Fc Fragment Specific; Anti-Human IgG<sub>2</sub>, Fd Fragment Specific; Anti-Human IgG<sub>3</sub>, Hinge Specific; Anti-Human IgG<sub>4</sub>, Fc Fragment Specific; Anti-Human IgM, Fc Fragment Specific; Anti-Human IgM,  $\mu$ -Chain Specific; Anti-Mouse IgE,  $\epsilon$ -Chain Specific; Anti-Mouse  $\gamma$ -Globulin (Goat); Anti-Mouse IgG,  $\gamma$ -Chain Specific (Goat); Anti-Mouse IgG,  $\gamma$ -Chain Specific (Goat) F(ab')<sub>2</sub> Fragment; Anti-Mouse IgG, H & L Chain Specific (Goat); Anti-Mouse IgM,  $\mu$ -Chain Specific (Goat); Anti-Mouse IgM, H & L Chain Specific (Goat); Anti-Rabbit  $\gamma$ -Globulin (Goat); Anti-Rabbit IgG, Fc Fragment Specific (Goat); Anti-Rabbit IgG, H & L Chain Specific (Goat); Anti-Rat  $\gamma$ -Globulin (Goat); Anti-Rat IgG, H & L Chain Specific; Anti-Rhesus Monkey  $\gamma$ -Globulin (Goat); and, Anti-Sheep IgG, H & L Chain Specific.

Another non-limiting list of the antibodies that may be produced using the present invention is provided in product catalogs of companies such as Phoenix Pharmaceuticals, Inc. ([www.phoenixpeptide.com](http://www.phoenixpeptide.com); 530 Harbor Boulevard, Belmont, CA), Peninsula Labs (San Carlos CA), SIGMA (St. Louis, MO [www.sigma-aldrich.com](http://www.sigma-aldrich.com)), Cappel ICN (Irvine, California, [www.icnbiomed.com](http://www.icnbiomed.com)), and Calbiochem (La Jolla, California, [www.calbiochem.com](http://www.calbiochem.com)), which are all incorporated herein by reference in their entirety. The polynucleotide sequences encoding these antibodies may be obtained from the scientific literature, from patents, and from databases such as GenBank. Alternatively, one of ordinary skill in the art may design the polynucleotide sequence to be incorporated into the genome by choosing the codons

that encode for each amino acid in the desired antibody. Antibodies made by the transgenic animals of the present invention include antibodies that may be used as therapeutic reagents, for example in cancer immunotherapy against specific antigens. Some of these antibodies include, but are not limited to, antibodies which bind the

5 following ligands: adrenomedulin, amylin, calcitonin, amyloid, calcitonin gene-related peptide, cholecystokinin, gastrin, gastric inhibitory peptide, gastrin releasing peptide, interleukin, interferon, cortistatin, somatostatin, endothelin, sarafotoxin, glucagon, glucagon-like peptide, insulin, atrial natriuretic peptide, BNP, CNP, neurokinin, substance P, leptin, neuropeptide Y, melanin concentrating hormone,

10 melanocyte stimulating hormone, orphanin, endorphin, dynorphin, enkephalin, enkephalin, leumorphin, peptide F, PACAP, PACAP-related peptide, parathyroid hormone, urocortin, corticotrophin releasing hormone, PHM, PHI, vasoactive intestinal polypeptide, secretin, ACTH, angiotensin, angiotensin, bombesin, endostatin, bradykinin, FMRF amide, galanin, gonadotropin releasing hormone

15 (GnRH) associated peptide, GnRH, growth hormone releasing hormone, inhibin, granulocyte-macrophage colony stimulating factor (GM-CSF), motilin, neurotensin, oxytocin, vasopressin, osteocalcin, pancreastatin, pancreatic polypeptide, peptide YY, proopiomelanocortin, transforming growth factor, vascular endothelial growth factor, vesicular monoamine transporter, vesicular acetylcholine transporter, ghrelin, NPW,

20 NPB, C3d, prokineticin, thyroid stimulating hormone, luteinizing hormone, follicle stimulating hormone, prolactin, growth hormone, beta-lipotropin, melatonin, kallikriens, kinins, prostaglandins, erythropoietin, p146 (SEQ ID NO:30 amino acid sequence, SEQ ID NO:31, nucleotide sequence), estrogen, testosterone, corticosteroids, mineralocorticoids, thyroid hormone, thymic hormones, connective

25 tissue proteins, nuclear proteins, actin, avidin, activin, agrin, albumin, and prohormones, propeptides, splice variants, fragments and analogs thereof.

The following is yet another non-limiting list of antibodies that can be produced by the methods of present invention: abciximab (ReoPro), abciximab anti-platelet aggregation monoclonal antibody, anti-CD11a (hu1124), anti-CD18 antibody,

30 anti-CD20 antibody, anti-cytomegalovirus (CMV) antibody, anti-digoxin antibody, anti-hepatitis B antibody, anti-HER-2 antibody, anti-idiotypic antibody to GD3 glycolipid, anti-IgE antibody, anti-IL-2R antibody, antimetastatic cancer antibody (mAb 17-1A), anti-rabies antibody, anti-respiratory syncytial virus (RSV) antibody, anti-Rh antibody, anti-TCR, anti-TNF antibody, anti-VEGF antibody and Fab

fragment thereof, rattlesnake venom antibody, black widow spider venom antibody, coral snake venom antibody, antibody against very late antigen-4 (VLA-4), C225 humanized antibody to EGF receptor, chimeric (human & mouse) antibody against TNF $\alpha$ , antibody directed against GPIIb/IIIa receptor on human platelets, gamma  
5 globulin, anti-hepatitis B immunoglobulin, human anti-D immunoglobulin, human antibodies against *S aureus*, human tetanus immunoglobulin, humanized antibody against the epidermal growth receptor-2, humanized antibody against the  $\alpha$  subunit of the interleukin-2 receptor, humanized antibody CTLA4IG, humanized antibody to the IL-2 R  $\alpha$ -chain, humanized anti-CD40-ligand monoclonal antibody (5c8), humanized  
10 mAb against the epidermal growth receptor-2, humanized mAb to rous sarcoma virus, humanized recombinant antibody (IgG1k) against respiratory syncytial virus (RSV), lymphocyte immunoglobulin (anti-thymocyte antibody), lymphocyte immunoglobulin, mAb against factor VII, MDX-210 bi-specific antibody against HER-2, MDX-22, MDX-220 bi-specific antibody against TAG-72 on tumors, MDX-  
15 33 antibody to Fc $\gamma$ R1 receptor, MDX-447 bi-specific antibody against EGF receptor, MDX-447 bispecific humanized antibody to EGF receptor, MDX-RA immunotoxin (ricin A linked) antibody, Medi-507 antibody (humanized form of BTI-322) against CD2 receptor on T-cells, monoclonal antibody LDP-02, muromonab-CD3(OKT3) antibody, OKT3 ("muromomab-CD3") antibody, PRO 542 antibody, ReoPro  
20 ("abciximab") antibody, and TNF-IgG fusion protein. It is to be understood that wherever the term "humanized" appears in the present patent application with regard to an antibody or molecule, that an antibody or molecule may be designed to be specific for any animal using selected polynucleotide sequences in the gene of interest included in the transposon-based vectors. Antibodies may be made against any  
25 selected antigen known to one of ordinary skill in the art.

The antibodies prepared using the methods of the present invention may also be designed to possess specific labels that may be detected through means known to one of ordinary skill in the art so that their location and distribution can be assessed following gene therapy and expression of the antibodies. The antibodies may also be  
30 designed to possess specific sequences useful for purification through means known to one of ordinary skill in the art. Specialty antibodies designed for binding specific antigens may also be made in transgenic animals using the transposon-based vectors of the present invention.

Production of a monoclonal antibody using the transposon-based vectors of the present invention can be accomplished in a variety of ways. In one embodiment, two vectors may be constructed: one that encodes the light chain, and a second vector that encodes the heavy chain of the monoclonal antibody. These vectors may then be  
5 incorporated into the genome of the target animal by methods disclosed herein. In an alternative embodiment, the sequences encoding light and heavy chains of a monoclonal antibody may be included on a single DNA construct. For example, the coding sequence of light and heavy chains of a murine monoclonal antibody that show specificity for human seminoprotein can be expressed using transposon-based  
10 constructs of the present invention (GenBank Accession numbers AY129006 and AY129304 for the light and heavy chains, respectively).

The transposon based vectors may include genes encoding proteins and peptides synthesized by the immune system including those synthesized by the thymus, lymph nodes, spleen, and the gastrointestinal associated lymph tissues  
15 (GALT) system. The immune system proteins and peptides proteins that can be made in transgenic animals using the transposon-based vectors of the present invention include, but are not limited to, alpha-interferon, beta-interferon, gamma-interferon, alpha-interferon A, alpha-interferon 1, G-CSF, GM-CSF, interleukin-1 (IL-1), IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, TNF- $\alpha$ , and TNF-  
20  $\beta$ . Other cytokines included in the present invention include cardiotrophin, stromal cell derived factors including stromal cell derived factor alpha, macrophage derived chemokine (MDC), melanoma growth stimulatory activity (MGSA), macrophage inflammatory proteins 1 alpha (MIP-1 alpha), 2, 3 alpha, 3 beta, 4 and 5, heat shock proteins (HSP) of different molecular weights (HSP-70, HSP-80, HSP-90 and others).  
25 Cell repellent molecules may also be made using the present invention, such as interleukins, stromal cell derived factor alpha and HSPs.

Lytic peptides, such as p146, are also included in the desired molecules that may be produced using the vectors and methods of the present invention. Lytic peptides are known to one of ordinary skill in the art and may be administered for  
30 gene therapy, for example to lyse cancer cells. In one embodiment, the p146 peptide comprises an amino acid sequence of SEQ ID NO:30. The present invention also encompasses a transposon-based vector comprising a p146 nucleic acid comprising a polynucleotide sequence of SEQ ID NO:31. Other lytic peptides and the class of



proteins called lysins may be made with the transposon-based vectors of the present invention.

Enzymes are another class of proteins that may be made through gene therapy of the transposon-based vectors of the present invention. Such enzymes include but  
5 are not limited to adenosine deaminase, alpha-galactosidase, cellulase, collagenase, dnaseI, hyaluronidase, lactase, L-asparaginase, pancreatin, papain, streptokinase B, subtilisin, superoxide dismutase, thrombin, trypsin, urokinase, fibrinolysin, glucocerebrosidase and plasminogen activator. Many diseases, such as genetic diseases, involve problems in the production of enzymes. Through the practice of the  
10 present invention, administration of the transposon based vectors encoding specific enzymes provides gene therapy to the animal or human. Examples of such conditions are known to one of ordinary skill in the art and include phenylketonuria, Tay-Sachs disease, and severe combined immunodeficiency disease, associated respectively with phenylalanine hydroxylase, hexosaminidase, and adenine deaminase. Other genetic  
15 disorders are described in Robbins Pathologic Basis of Disease, Cotran et al. eds. 6<sup>th</sup> ed., pp 139-187, 1999 Saunders, and in Harrison's Principles of Internal Medicine, Fauci et al. eds. 14<sup>th</sup> ed. pp. 365-409, 1998, McGraw Hill. In some embodiments wherein the enzyme could have deleterious effects, additional amino acids and a protease cleavage site are added to the carboxy end of the enzyme of interest in order  
20 to prevent expression of a functional enzyme. Subsequent digestion of the enzyme with a protease results in activation of the enzyme.

Extracellular matrix proteins are one class of desired proteins that may be made through the gene therapy methods of the present invention. Examples include  
25 but are not limited to collagen, fibrin, elastin, laminin, and fibronectin and subtypes thereof. Animals receiving gene therapy for conditions such as arthritis or clotting disorders may make some of these matrix proteins. Gene therapy may be administered to stimulate formation of cartilage, such as articular cartilage, or for deposition of new bone. Intracellular proteins and structural proteins are other classes of desired proteins in the present invention.

30 Growth factors are another desired class of proteins that may be made through the gene therapy methods of the present invention and include, but are not limited to, transforming growth factor- $\alpha$  ("TGF- $\alpha$ "), transforming growth factor- $\beta$  (TGF- $\beta$ ), platelet-derived growth factors (PDGF), fibroblast growth factors (FGF), including

FGF acidic isoforms 1 and 2, FGF basic form 2 and FGF 4, 8, 9 and 10, nerve growth factors (NGF) including NGF 2.5s, NGF 7.0s and beta NGF and neurotrophins, brain derived neurotrophic factor, cartilage derived factor, growth factors for stimulation of the production of red blood cells, growth factors for stimulation of the production of  
5 white blood cells, bone growth factors (BGF), basic fibroblast growth factor, vascular endothelial growth factor (VEGF), granulocyte colony stimulating factor (G-CSF), insulin like growth factor (IGF) I and II, hepatocyte growth factor, glial neurotrophic growth factor (GDNF), stem cell factor (SCF), keratinocyte growth factor (KGF), transforming growth factors (TGF), including TGFs alpha, beta, beta1, beta2, beta3,  
10 skeletal growth factor, bone matrix derived growth factors, bone derived growth factors, erythropoietin (EPO) and mixtures thereof.

Another desired class of proteins that may be made may be made through the gene therapy of the present invention include, but are not limited to, leptin, leukemia inhibitory factor (LIF), tumor necrosis factor alpha and beta, ENBREL, angiostatin,  
15 endostatin, thrombospondin, osteogenic protein-1, bone morphogenetic proteins 2 and 7, osteonectin, somatomedin-like peptide, and osteocalcin.

Yet another desired class of proteins are blood proteins or clotting cascade protein including albumin, Prekallikrein, High molecular weight kininogen (HMWK) (contact activation cofactor; Fitzgerald, Flaujeac Williams factor), Factor I  
20 (Fibrinogen), Factor II (prothrombin), Factor III (Tissue Factor), Factor IV (calcium), Factor V (proaccelerin, labile factor, accelerator (Ac-) globulin), Factor VI (Va) (accelerin), Factor VII (proconvertin), serum prothrombin conversion accelerator (SPCA), cothromboplastin), Factor VIII (antihemophilic factor A, antihemophilic globulin (AHG)), Factor IX (Christmas Factor, antihemophilic factor B, plasma  
25 thromboplastin component (PTC)), Factor X (Stuart-Prower Factor), Factor XI (Plasma thromboplastin antecedent (PTA)), Factor XII (Hageman Factor), Factor XIII (protransglutaminase, fibrin stabilizing factor (FSF), fibrinoligase), von Willibrand factor, Protein C, Protein S, Thrombomodulin, Antithrombin III.

A non-limiting list of the peptides and proteins that may be made may be  
30 made through the use of the gene therapy methods of the present invention is provided in product catalogs of companies such as Phoenix Pharmaceuticals, Inc. ([www.phoenixpeptide.com](http://www.phoenixpeptide.com); 530 Harbor Boulevard, Belmont, CA), Peninsula Labs (San Carlos CA), SIGMA, (St. Louis, MO [www.sigma-aldrich.com](http://www.sigma-aldrich.com)), Cappel ICN (Irvine, California, [www.icnbiomed.com](http://www.icnbiomed.com)), and Calbiochem (La Jolla, California,

www.calbiochem.com). The polynucleotide sequences encoding these proteins and peptides of interest may be obtained from the scientific literature, from patents, and from databases, such as GenBank. Alternatively, one of ordinary skill in the art may design the polynucleotide sequence to be incorporated into the genome by choosing  
5 the codons that encode for each amino acid in the desired protein or peptide.

Some of these desired proteins or peptides that may be made through the use of the gene therapy methods of the present invention include but are not limited to the following: adrenomedulin, amylin, calcitonin, amyloid, calcitonin gene-related peptide, cholecystokinin, gastrin, gastric inhibitory peptide, gastrin releasing peptide,  
10 interleukin, interferon, cortistatin, somatostatin, endothelin, sarafotoxin, glucagon, glucagon-like peptide, insulin, atrial natriuretic peptide, BNP, CNP, neurokinin, substance P, leptin, neuropeptide Y, melanin concentrating hormone, melanocyte stimulating hormone, orphanin, endorphin, dynorphin, enkephalin, leumorphin, peptide F, PACAP, PACAP-related peptide, parathyroid hormone, urocortin,  
15 corticotrophin releasing hormone, PHM, PHI, vasoactive intestinal polypeptide, secretin, ACTH, angiotensin, angiostatin, bombesin, endostatin, bradykinin, FMRF amide, galanin, gonadotropin releasing hormone (GnRH) associated peptide, GnRH, growth hormone releasing hormone, inhibin, granulocyte-macrophage colony stimulating factor (GM-CSF), motilin, neurotensin, oxytocin, vasopressin,  
20 osteocalcin, pancreastatin, pancreatic polypeptide, peptide YY, proopiomelanocortin, transforming growth factor, vascular endothelial growth factor, vesicular monoamine transporter, vesicular acetylcholine transporter, ghrelin, NPW, NPB, C3d, prokineticin, thyroid stimulating hormone, luteinizing hormone, follicle stimulating hormone, prolactin, growth hormone, beta-lipotropin, melatonin, kallikriens, kinins,  
25 prostaglandins, erythropoietin, p146 (SEQ ID NO:30, amino acid sequence, SEQ ID NO:31, nucleotide sequence), thymic hormones, connective tissue proteins, nuclear proteins, actin, avidin, activin, agrin, albumin, apolipoproteins, apolipoprotein A, apolipoprotein B, and prohormones, propeptides, splice variants, fragments and analogs thereof.

30 Other desired proteins that may be made by the transgenic animals receiving gene therapy according to the present invention include bacitracin, polymixin b, vancomycin, cyclosporine, anti-RSV antibody, alpha-1 antitrypsin (AAT), anti-cytomegalovirus antibody, anti-hepatitis antibody, anti-inhibitor coagulant complex, anti-rabies antibody, anti-Rh(D) antibody, adenosine deaminase, anti-digoxin

antibody, antivenin crotalidae (rattlesnake venom antibody), antivenin latrodectus (black widow spider venom antibody), antivenin micrurus (coral snake venom antibody), aprotinin, corticotropin (ACTH), diphtheria antitoxin, lymphocyte immune globulin (anti-thymocyte antibody), protamine, thyrotropin, capreomycin,  $\alpha$ -galactosidase, gramicidin, streptokinase, tetanus toxoid, tyrothricin, IGF-1, proteins of

5 galactosidase, gramicidin, streptokinase, tetanus toxoid, tyrothricin, IGF-1, proteins of varicella vaccine, anti-TNF antibody, anti-IL-2r antibody, anti-HER-2 antibody, OKT3 ("muromonab-CD3") antibody, TNF-IgG fusion protein, ReoPro ("abciximab") antibody, ACTH fragment 1-24, desmopressin, gonadotropin-releasing hormone, histrelin, leuprolide, lypressin, nafarelin, peptide that binds GPIIb/GPIIIa on

10 platelets (integrilin), goserelin, capreomycin, colistin, anti-respiratory syncytial virus, lymphocyte immune globulin (Thymoglovin, Atgam), panorex, alpha-antitrypsin, botulinum, lung surfactant protein, tumor necrosis receptor-IgG fusion protein (enbrel), gonadorelin, proteins of influenza vaccine, proteins of rotavirus vaccine, proteins of haemophilus b conjugate vaccine, proteins of poliovirus vaccine, proteins of

15 pneumococcal conjugate vaccine, proteins of meningococcal C vaccine, proteins of influenza vaccine, megakaryocyte growth and development factor (MGDF), neuroimmunophilin ligand-A (NIL-A), brain-derived neurotrophic factor (BDNF), glial cell line-derived neurotrophic factor (GDNF), leptin (native), leptin B, leptin C, IL-1RA (interleukin-1RA), R-568, novel erythropoiesis-stimulating protein (NESP),

20 humanized mAb to rous sarcoma virus (MEDI-493), glutamyl-tryptophan dipeptide IM862, LFA-3TIP immunosuppressive, humanized anti-CD40-ligand monoclonal antibody (5c8), gelsolin enzyme, tissue factor pathway inhibitor (TFPI), proteins of meningitis B vaccine, antimetastatic cancer antibody (mAb 17-1A), chimeric (human & mouse) mAb against TNF $\alpha$ , mAb against factor VII, relaxin, capreomycin,

25 glycopeptide (LY333328), recombinant human activated protein C (rhAPC), humanized mAb against the epidermal growth receptor-2, altepase, anti-CD20 antigen, C2B8 antibody, insulin-like growth factor-1, atrial natriuretic peptide (anaritide), tenectapase, anti-CD11a antibody (hu 1124), anti-CD18 antibody, mAb LDP-02, anti-VEGF antibody, Fab fragment of anti-VEGF Ab, APO2 ligand (tumor

30 necrosis factor-related apoptosis-inducing ligand), rTGF- $\beta$  (transforming growth factor- $\beta$ ), alpha-antitrypsin, ananain (a pineapple enzyme), humanized mAb CTLA4IG, PRO 542 (mAb), D2E7 (mAb), calf intestine alkaline phosphatase,  $\alpha$ -L-iduronidase,  $\alpha$ -L-galactosidase (human glutamic acid decarboxylase, acid

sphingomyelinase, bone morphogenetic protein-2 (rhBMP-2), proteins of HIV vaccine, T cell receptor (TCR) peptide vaccine, TCR peptides, V beta 3 and V beta 13.1. (IR502), (IR501), BI 1050/1272 mAb against very late antigen-4 (VLA-4), C225 humanized mAb to EGF receptor, anti-idiotypic antibody to GD3 glycolipid, antibacterial peptide against *H. pylori*, MDX-447 bispecific humanized mAb to EGF receptor, anti-cytomegalovirus (CMV), Medi-491 B19 parvovirus vaccine, humanized recombinant mAb (IgG1k) against respiratory syncytial virus (RSV), urinary tract infection vaccine (against "pili" on *Escherichia coli* strains), proteins of lyme disease vaccine against *B. burgdorferi* protein (DbpA), proteins of Medi-501 human papilloma virus-11 vaccine (HPV), *Streptococcus pneumoniae* vaccine, Medi-507 mAb (humanized form of BTI-322) against CD2 receptor on T-cells, MDX-33 mAb to FcγR1 receptor, MDX-RA immunotoxin (ricin A linked) mAb, MDX-210 bi-specific mAb against HER-2, MDX-447 bi-specific mAb against EGF receptor, MDX-22, MDX-220 bi-specific mAb against TAG-72 on tumors, colony-stimulating factor (CSF) (molgramostim), humanized mAb to the IL-2 R α-chain (basiliximab), mAb to IgE (IGE 025A), myelin basic protein-altered peptide (MSP771A), humanized mAb against the epidermal growth receptor-2, humanized mAb against the α subunit of the interleukin-2 receptor, low molecular weight heparin, anti-hemophilic factor, and bactericidal/permeability-increasing protein (r-BPI).

The peptides and proteins made by animals receiving gene therapy using the present invention may be labeled using labels and techniques known to one of ordinary skill in the art. Some of these labels are described in the "Handbook of Fluorescent Probes and Research Products", ninth edition, Richard P. Haugland (ed) Molecular Probes, Inc. Eugene, OR), which is incorporated herein in its entirety. Some of these labels may be genetically engineered into the polynucleotide sequence for the expression of the selected protein or peptide. The peptides and proteins may also have label-incorporation "handles" incorporated to allow labeling of an otherwise difficult or impossible to label protein.

It is to be understood that the various classes of desired peptides and proteins, as well as specific peptides and proteins described in this section may be modified as described below by inserting selected codons for desired amino acid substitutions into the gene incorporated into the transgenic animal.

#### Genes of Interest Encoding Desired Nucleic Acids and Other Molecule

The present invention may also be used to produce desired molecules other than proteins and peptides including, but not limited to, lipoproteins such as high density lipoprotein (HDL), HDL-Milano, and low density lipoprotein, lipids, carbohydrates, siRNA and ribozymes. In these embodiments, a gene of interest  
5 encodes a nucleic acid molecule or a protein that directs production of the desired molecule.

#### Nucleic Acids

RNAi technology can be directed against numerous aberrant genes, including those that allow proliferation of tumor cells. A variety of strategies can be used to  
10 inhibit cancer. These include the inhibition of overexpressed oncogenes, blocking cell division by interfering with cyclin E and related genes or promoting apoptosis by suppressing antiapoptotic genes. RNAi against multidrug resistance genes or chemoresistance targets may also provide useful cancer treatments. A non-limiting list of gene and protein targets for cancer therapy is found in Table 2 (M. Izquierdo.  
15 2004. Short interfering RNAs as a tool for cancer gene therapy. Cancer Gene Therapy pp 1-11).

There are guidelines for designing dsRNA for use as RNAi therapy. These rules are known to one of ordinary skill in the art. Generally, the dsRNA cannot be shorter than 21 nucleotides (nt) or longer than 30 nt so the antiviral interferon  
20 response is not triggered. Other features to be avoided include, tight stem loops, inverted repeats, high sequence homology with other genes, and a lack of 4 or more consecutive T or A to avoid premature pol III transcription termination. Features to include are: 1) Initiation with a G or C after an AA in the 5' flanking sequence; 2) sense strand base preferences at positions 3 (A), 10 (U), 13 (A), and 19 (A); and 3)  
25 low G/C content (30-60%) (M. Izquierdo. 2004. Short interfering RNAs as a tool for cancer gene therapy. Cancer Gene Therapy pp 1-11).

The present invention provides a new and effective method for delivering shRNA using transposon-based vectors. This method can be used to treat various conditions and diseases and is a method of providing gene therapy. shRNA would be  
30 administered using a transposon based vector targeted to a specific cell type. The transposase (ATS) can be expressed by a cell-specific promoter (Table 1) to limit incorporation into a specific cell, and/or a cell-specific promoter could be used to express an shRNA to a gene listed in this document or any gene to be targeted for inactivation. In addition to the genes listed as targets for cancer therapy (see also

Table 3), genes such as apoB (apolipoprotein B) to lower cholesterol (Akinc, et al. 2004. *Nature* 432(7017): 155-156), viral genes to eliminate hepatitis B and C (Shlomai and Shaul. 2004. *Liver Int* 6:526-31), metabolism and obesity genes (Campion, et al. 2004 *Nutr. Rev* 62:321-330), HIV (Berkhout. 2004. *Curr Opin Mol Ther* 6:141-145; Takaku. 2004. *Antivir Chem Chemother*. 15:57-65), cardiac disease through down regulation of phospholamban (PL; Poller et al. 2004. *Z Kardiol* 93: 171-193), and 5' nontranslated region (5' NTR) of hepatitis C (Kronke et al. 2004. *J Virol*. 78:3436-3446).

The present invention further encompasses gene therapy to produce inhibitory molecules to inhibit endogenous (i.e., non-vector) protein production. Such therapy may be used to inhibit a gene that is over expressed. These inhibitory molecules include antisense nucleic acids, siRNA, polynucleotide strands that affect cellular function and inhibitory proteins. In one embodiment, the endogenous protein whose expression is inhibited is an egg white protein including, but not limited to ovalbumin, ovotransferrin, ovomucin, ovomucin inhibitor, cystatin, ovostatin, lysozyme, ovoglobulin G2, ovoglobulin G3, avidin, or thiamin binding protein. In one embodiment, a transposon-based vector containing an ovalbumin DNA sequence, that upon transcription forms a double stranded RNA molecule, is transfected into an animal, such as a bird, and the bird's production of endogenous ovalbumin protein is reduced by the interference RNA mechanism (RNAi). In other embodiments, a transposon-based vector encodes an inhibitory RNA molecule that inhibits the expression of more than one egg white protein. One exemplary construct is provided in Figure 4 wherein "Ovgen" indicates approximately 60 base pairs of an ovalbumin gene, "Ovotrans" indicates approximately 60 base pairs of an ovotransferrin gene and "Ovomucin" indicates approximately 60 base pairs of an ovomucin gene. These ovalbumin, ovotransferrin and ovomucin can be from any avian species, and in some embodiments, are from a chicken or quail. The term "pro" indicates the pro portion of a prepro sequence. One exemplary prepro sequence is that of cecropin and comprising base pairs 563-733 of the Cecropin cap site and Prepro provided in Genbank accession number X07404.

Additionally, inducible knockouts or knockdowns of the endogenous protein may be created to achieve a reduction or inhibition of endogenous protein production. The approach may be used for inhibition of any selected endogenous protein in animals receiving gene therapy.

Modified Desired Proteins and Peptides Made by Animals Receiving Gene Therapy

"Proteins", "peptides", "polypeptides" and "oligopeptides" are chains of amino acids (typically L-amino acids) whose alpha carbons are linked through peptide bonds  
5 formed by a condensation reaction between the carboxyl group of the alpha carbon of one amino acid and the amino group of the alpha carbon of another amino acid. The terminal amino acid at one end of the chain (i.e., the amino terminal) has a free amino group, while the terminal amino acid at the other end of the chain (i.e., the carboxy terminal) has a free carboxyl group. As such, the term "amino terminus" (abbreviated  
10 N-terminus) refers to the free alpha-amino group on the amino acid at the amino terminal of the protein, or to the alpha-amino group (imino group when participating in a peptide bond) of an amino acid at any other location within the protein. Similarly, the term "carboxy terminus" (abbreviated C-terminus) refers to the free carboxyl group on the amino acid at the carboxy terminus of a protein, or to the  
15 carboxyl group of an amino acid at any other location within the protein.

Typically, the amino acids making up a protein are numbered in order, starting at the amino terminal and increasing in the direction toward the carboxy terminal of the protein. Thus, when one amino acid is said to "follow" another, that amino acid is positioned closer to the carboxy terminal of the protein than the preceding amino acid.

20 The term "residue" is used herein to refer to an amino acid (D or L) or an amino acid mimetic that is incorporated into a protein by an amide bond. As such, the amino acid may be a naturally occurring amino acid or, unless otherwise limited, may encompass known analogs of natural amino acids that function in a manner similar to the naturally occurring amino acids (i.e., amino acid mimetics). Moreover, an amide  
25 bond mimetic includes peptide backbone modifications well known to those skilled in the art.

Furthermore, one of skill will recognize that, as mentioned above, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than about 5%, more typically less  
30 than about 1%) in an encoded sequence are conservatively modified variations where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Such substitutions may be engineered by selecting the desired nucleotides for insertion into the gene of interest in animals receiving gene therapy. Conservative substitutions in polynucleotide sequences are included within the scope



of the present invention, wherein codons in a sequence may be replaced with other codons encoding for conservatively substituted amino acids, as explained below in the conservative substitution table. In other words, a codon in a polynucleotide sequence encoding for an alanine may be substituted with a codon encoding for a valine.

5 Conservative substitution tables providing functionally similar amino acids are well known in the art. The following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
- 2) Aspartic acid (D), Glutamic acid (E);
- 10 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

A conservative substitution is a substitution in which the substituting amino acid (naturally occurring or modified) is structurally related to the amino acid being substituted, i.e., has about the same size and electronic properties as the amino acid being substituted. Thus, the substituting amino acid would have the same or a similar functional group in the side chain as the original amino acid. A "conservative substitution" also refers to utilizing a substituting amino acid which is identical to the amino acid being substituted except that a functional group in the side chain is protected with a suitable protecting group.

Suitable protecting groups are described in Green and Wuts, "Protecting Groups in Organic Synthesis", John Wiley and Sons, Chapters 5 and 7, 1991, the teachings of which are incorporated herein by reference. Preferred protecting groups are those which facilitate transport of the peptide through membranes, for example, by reducing the hydrophilicity and increasing the lipophilicity of the peptide, and which can be cleaved, either by hydrolysis or enzymatically (Ditter et al., 1968. J. Pharm. Sci. 57:783; Ditter et al., 1968. J. Pharm. Sci. 57:828; Ditter et al., 1969. J. Pharm. Sci. 58:557; King et al., 1987. Biochemistry 26:2294; Lindberg et al., 1989. Drug Metabolism and Disposition 17:311; Tunek et al., 1988. Biochem. Pharm. 37:3867; Anderson et al., 1985 Arch. Biochem. Biophys. 239:538; and Singhal et al., 1987. FASEB J. 1:220). Suitable hydroxyl protecting groups include ester, carbonate and carbamate protecting groups. Suitable amine protecting groups include acyl groups and alkoxy or aryloxy carbonyl groups, as described above for N-terminal protecting

groups. Suitable carboxylic acid protecting groups include aliphatic, benzyl and aryl esters, as described below for C-terminal protecting groups. In one embodiment, the carboxylic acid group in the side chain of one or more glutamic acid or aspartic acid residues in a peptide of the present invention is protected, preferably as a methyl, ethyl, benzyl or substituted benzyl ester, more preferably as a benzyl ester.

Provided below are groups of naturally occurring and modified amino acids in which each amino acid in a group has similar electronic and steric properties. Thus, a conservative substitution can be made by substituting an amino acid with another amino acid from the same group. Such substitutions may be engineered through selection of the appropriate nucleotides in constructing the gene of interest for introduction into animals receiving gene therapy. It is to be understood that these groups are non-limiting, i.e. that there are additional modified amino acids which could be included in each group.

Group I includes leucine, isoleucine, valine, methionine and modified amino acids having the following side chains: ethyl, n-propyl n-butyl. Preferably, Group I includes leucine, isoleucine, valine and methionine.

Group II includes glycine, alanine, valine and a modified amino acid having an ethyl side chain. Preferably, Group II includes glycine and alanine.

Group III includes phenylalanine, phenylglycine, tyrosine, tryptophan, cyclohexylmethyl glycine, and modified amino residues having substituted benzyl or phenyl side chains. Preferred substituents include one or more of the following: halogen, methyl, ethyl, nitro,  $\text{—NH}_2$ , methoxy, ethoxy and  $\text{—CN}$ . Preferably, Group III includes phenylalanine, tyrosine and tryptophan.

Group IV includes glutamic acid, aspartic acid, a substituted or unsubstituted aliphatic, aromatic or benzylic ester of glutamic or aspartic acid (e.g., methyl, ethyl, n-propyl iso-propyl, cyclohexyl, benzyl or substituted benzyl), glutamine, asparagine,  $\text{—CO—NH—}$  alkylated glutamine or asparagines (e.g., methyl, ethyl, n-propyl and iso-propyl) and modified amino acids having the side chain  $\text{—(CH}_2)_3\text{—COOH}$ , an ester thereof (substituted or unsubstituted aliphatic, aromatic or benzylic ester), an amide thereof and a substituted or unsubstituted N-alkylated amide thereof. Preferably, Group IV includes glutamic acid, aspartic acid, methyl aspartate, ethyl aspartate, benzyl aspartate and methyl glutamate, ethyl glutamate and benzyl glutamate, glutamine and asparagine.

Group V includes histidine, lysine, ornithine, arginine, N-nitroarginine,  $\beta$ -cycloarginine,  $\gamma$ -hydroxyarginine, N-amidinocitruline and 2-amino-4-guanidinobutanoic acid, homologs of lysine, homologs of arginine and homologs of ornithine. Preferably, Group V includes histidine, lysine, arginine and ornithine. A homolog of an amino acid includes from 1 to about 3 additional or subtracted methylene units in the side chain.

Group VI includes serine, threonine, cysteine and modified amino acids having C1-C5 straight or branched alkyl side chains substituted with  $-\text{OH}$  or  $-\text{SH}$ , for example,  $-\text{CH}_2\text{CH}_2\text{OH}$ ,  $-\text{CH}_2\text{CH}_2\text{CH}_2\text{OH}$  or  $-\text{CH}_2\text{CH}_2\text{OHCH}_3$ . Preferably, Group VI includes serine, cysteine or threonine.

In another aspect, suitable substitutions for amino acid residues include "severe" substitutions. A "severe substitution" is a substitution in which the substituting amino acid (naturally occurring or modified) has significantly different size and/or electronic properties compared with the amino acid being substituted. Thus, the side chain of the substituting amino acid can be significantly larger (or smaller) than the side chain of the amino acid being substituted and/or can have functional groups with significantly different electronic properties than the amino acid being substituted. Examples of severe substitutions of this type include the substitution of phenylalanine or cyclohexylmethyl glycine for alanine, isoleucine for glycine, a D amino acid for the corresponding L amino acid, or  $-\text{NH}-\text{CH}[(\text{CH}_2)_5-\text{COOH}]-\text{CO}-$  for aspartic acid. Alternatively, a functional group may be added to the side chain, deleted from the side chain or exchanged with another functional group. Examples of severe substitutions of this type include adding of valine, leucine or isoleucine, exchanging the carboxylic acid in the side chain of aspartic acid or glutamic acid with an amine, or deleting the amine group in the side chain of lysine or ornithine. In yet another alternative, the side chain of the substituting amino acid can have significantly different steric and electronic properties than the functional group of the amino acid being substituted. Examples of such modifications include tryptophan for glycine, lysine for aspartic acid and  $(\text{CH}_2)_4\text{COOH}$  for the side chain of serine. These examples are not meant to be limiting.

In another embodiment, for example in the synthesis of a peptide 26 amino acids in length, the individual amino acids may be substituted according in the following manner:

- AA<sub>1</sub> is serine, glycine, alanine, cysteine or threonine;  
 AA<sub>2</sub> is alanine, threonine, glycine, cysteine or serine;  
 AA<sub>3</sub> is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-nitroarginine, β-cycloarginine, γ-hydroxyarginine, N-amidinocitruline or 2-amino-4-guanidinobutanoic acid;  
 5 AA<sub>4</sub> is proline, leucine, valine, isoleucine or methionine;  
 AA<sub>5</sub> is tryptophan, alanine, phenylalanine, tyrosine or glycine;  
 AA<sub>6</sub> is serine, glycine, alanine, cysteine or threonine;  
 AA<sub>7</sub> is proline, leucine, valine, isoleucine or methionine;  
 10 AA<sub>8</sub> is alanine, threonine, glycine, cysteine or serine;  
 AA<sub>9</sub> is alanine, threonine, glycine, cysteine or serine;  
 AA<sub>10</sub> is leucine, isoleucine, methionine or valine;  
 AA<sub>11</sub> is serine, glycine, alanine, cysteine or threonine;  
 AA<sub>12</sub> is leucine, isoleucine, methionine or valine;  
 15 AA<sub>13</sub> is leucine, isoleucine, methionine or valine;  
 AA<sub>14</sub> is glutamine, glutamic acid, aspartic acid, asparagine, or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;  
 AA<sub>15</sub> is arginine, N-nitroarginine, β-cycloarginine, γ-hydroxy-arginine, N-amidinocitruline or 2-amino-4-guanidino-butanoic acid  
 20 AA<sub>16</sub> is proline, leucine, valine, isoleucine or methionine;  
 AA<sub>17</sub> is serine, glycine, alanine, cysteine or threonine;  
 AA<sub>18</sub> is glutamic acid, aspartic acid, asparagine, glutamine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;  
 AA<sub>19</sub> is aspartic acid, asparagine, glutamic acid, glutamine, leucine, valine, isoleucine,  
 25 methionine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;  
 AA<sub>20</sub> is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-nitroarginine, β-cycloarginine, γ-hydroxyarginine, N-amidinocitruline or 2-amino-4-guanidinobutanoic acid;  
 30 AA<sub>21</sub> is alanine, threonine, glycine, cysteine or serine;  
 AA<sub>22</sub> is alanine, threonine, glycine, cysteine or serine;  
 AA<sub>23</sub> is histidine, serine, threonine, cysteine, lysine or ornithine;  
 AA<sub>24</sub> is threonine, aspartic acid, serine, glutamic acid or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;

AA<sub>25</sub> is asparagine, aspartic acid,, glutamic acid, glutamine, leucine, valine, isoleucine, methionine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid; and

AA<sub>26</sub> is cysteine, histidine, serine, threonine, lysine or ornithine.

5           It is to be understood that these amino acid substitutions may be made for longer or shorter peptides than the 26 mer in the preceding example above, and for proteins.

          In one embodiment of the present invention, codons for the first several N-terminal amino acids of the transposase are modified such that the third base of each  
10       codon is changed to an A or a T without changing the corresponding amino acid. It is preferable that between approximately 1 and 20, more preferably 3 and 15, and most preferably between 4 and 12 of the first N-terminal codons of the gene of interest are modified such that the third base of each codon is changed to an A or a T without changing the corresponding amino acid. In one embodiment, the first ten N-terminal  
15       codons of the gene of interest are modified in this manner.

          When several desired proteins, protein fragments or peptides are encoded in the gene of interest to be incorporated into the genome, one of skill in the art will appreciate that the proteins, protein fragments or peptides may be separated by a spacer molecule such as, for example, a peptide, consisting of one or more amino  
20       acids. Generally, the spacer will have no specific biological activity other than to join the desired proteins, protein fragments or peptides together, or to preserve some minimum distance or other spatial relationship between them. However, the constituent amino acids of the spacer may be selected to influence some property of the molecule such as the folding, net charge, or hydrophobicity. The spacer may also  
25       be contained within a nucleotide sequence with a purification handle or be flanked by cleavage sites, such as proteolytic cleavage sites.

          Such polypeptide spacers may have from about 1 to about 100 amino acids, preferably 3 to 20 amino acids, and more preferably 4-15 amino acids . The spacers in a polypeptide are independently chosen, but are preferably all the same. The  
30       spacers should allow for flexibility of movement in space and are therefore typically rich in small amino acids, for example, glycine, serine, proline or alanine. Preferably, peptide spacers contain at least 60%, more preferably at least 80% glycine or alanine. In addition, peptide spacers generally have little or no biological and antigenic activity. Preferred spacers are (Gly-Pro-Gly-Gly)<sub>x</sub> (SEQ ID NO:32) and (Gly<sub>4</sub>-Ser)<sub>y</sub>,

wherein x is an integer from about 3 to about 9 and y is an integer from about 1 to about 8. Specific examples of suitable spacers include

(Gly-Pro-Gly-Gly)<sub>3</sub>

SEQ ID NO:33 Gly Pro Gly Gly Gly Pro Gly Gly Gly Pro Gly Gly

5 (Gly<sub>4</sub>-Ser)<sub>3</sub>

SEQ ID NO:34 Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser

or (Gly<sub>4</sub>-Ser)<sub>4</sub>

SEQ ID NO:35 Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser  
Gly Gly Gly Gly Ser.

10 Nucleotide sequences encoding for the production of residues which may be useful in purification of the expressed recombinant protein in animals receiving gene therapy may also be built into the vector. Such sequences are known in the art and include the glutathione binding domain from glutathione S-transferase, polylysine, hexa-histidine or other cationic amino acids, thioredoxin, hemagglutinin antigen and  
15 maltose binding protein.

Additionally, nucleotide sequences may be inserted into the gene of interest to be incorporated so that the protein or peptide can also include from one to about six amino acids that create signals for proteolytic cleavage. In this manner, if a gene is designed to make one or more peptides or proteins of interest in the transgenic animal,  
20 specific nucleotide sequences encoding for amino acids recognized by enzymes may be incorporated into the gene to facilitate cleavage of the large protein or peptide sequence into desired peptides or proteins or both. For example, nucleotides encoding a proteolytic cleavage site can be introduced into the gene of interest so that a signal sequence can be cleaved from a protein or peptide encoded by the gene of interest.  
25 Nucleotide sequences encoding other amino acid sequences which display pH sensitivity or chemical sensitivity may also be added to the vector to facilitate separation of the signal sequence from the peptide or protein of interest.

Proteolytic cleavage sites include cleavage sites recognized by exopeptidases such as carboxypeptidase A, carboxypeptidase B, aminopeptidase I, and  
30 dipeptidylaminopeptidase; endopeptidases such as trypsin, V8-protease, enterokinase, factor Xa, collagenase, endoproteinase, subtilisin, and thrombin; and proteases such as Protease 3C IgA protease (Igase) Rhinovirus 3C(preScission)protease. Chemical cleavage sites are also included in the definition of cleavage site as used herein.

Chemical cleavage sites include, but are not limited to, site cleaved by cyanogen bromide, hydroxylamine, formic acid, and acetic acid.

In one embodiment of the present invention, a TAG sequence is linked to the gene of interest. The TAG sequence serves three purposes: 1) it allows free rotation  
5 of the peptide or protein to be isolated so there is no interference from the native protein or signal sequence, i.e. vitellogenin, 2) it provides a "purification handle" to isolate the protein using column purification, and 3) it includes a cleavage site to remove the desired protein from the signal and purification sequences. Accordingly, as used herein, a TAG sequence includes a spacer sequence, a purification handle and  
10 a cleavage site. The spacer sequences in the TAG proteins contain one or more repeats shown in SEQ ID NO:36. A preferred spacer sequence comprises the sequence provided in SEQ ID NO:37. One example of a purification handle is the gp41 hairpin loop from HIV I. Exemplary gp41 polynucleotide and polypeptide sequences are provided in SEQ ID NO:38 and SEQ ID NO:39, respectively.  
15 However, it should be understood that any antigenic region may be used as a purification handle, including any antigenic region of gp41. Preferred purification handles are those that elicit highly specific antibodies. Additionally, the cleavage site can be any protein cleavage site known to one of ordinary skill in the art and includes an enterokinase cleavage site comprising the Asp Asp Asp Asp Lys sequence (SEQ  
20 ID NO:40) and a furin cleavage site. Constructs containing a TAG sequence are shown in Figures 2 and 3. In one embodiment of the present invention, the TAG sequence comprises a polynucleotide sequence of SEQ ID NO:41.

#### Gene Therapy

25 Administration of the transposon based vectors of the present invention to achieve gene therapy in animals may be used to treat numerous genetic and non-genetic disorders.

DNA constructs of the present invention can be used to transform any animal cell, including but not limited to: cells producing hormones, cytokines, growth  
30 factors, or any other biologically active substance; cells of the immune system; cells of the nervous system; muscle (striatal, cardiac, smooth) cells; vascular system cells; endothelial cells; skin cells; mammary cells; and lung cells, including bronchial and alveolar cells. Transformation of any endocrine cell by a transposon-based DNA construct is contemplated as a part of a present invention. DNA constructs of the

present invention can be used to modulate, including both stimulation and inhibition, production of any substance, including but not limited to a hormone, a cytokine, or a growth factor, by an animal cell. Modulation of a regulated signal within a cell or a tissue, such as production of a second messenger, is also contemplated as a part of the present invention. In one aspect of the present invention, cells of the immune system may be the target for incorporation of a desired gene or genes encoding for production of antibodies. Accordingly, the thymus, bone marrow, beta lymphocytes (or B cells), gastrointestinal associated lymphatic tissue (GALT), Peyer's patches, bursa Fabricius, lymph nodes, spleen, and tonsil, and any other lymphatic tissue, may all be targets for administration of the compositions of the present invention. Use of the DNA constructs of the present invention is contemplated for treatment of any animal disease or condition that results from underproduction (such as diabetes) or overproduction (such as hyperthyroidism) of a hormone or other endogenous biologically active substance. Use of DNA constructs of the present invention to integrate nucleotide sequences encoding RNA molecules, such as anti-sense RNA or short interfering RNA, is also contemplated as a part of the present invention.

#### Genetic disorders

Genetic disorders are well known to one of ordinary skill in the art and may include, but are not limited to, general classes of mutations, Mendelian disorders, disorders with multifactorial inheritance, cytogenetic disorders, and single gene disorders with nonclassic inheritance. Many genetic disorders are described in Robbins Pathologic Basis of Disease, Cotran et al. eds. 6<sup>th</sup> ed., pp 139-187, 1999 Saunders, and in Harrison's Principles of Internal Medicine, Fauci et al. eds. 14<sup>th</sup> ed. pp. 365-409, 1998, McGraw Hill. Genetic disorders that may be treated with the method of the present invention include, but are not limited to those presented in Table 3, which also identifies the gene and often the chromosome associated with the specific genetic disorder.

Mendelian disorders include autosomal dominant disorders autosomal recessive disorders and X-linked disorders. Such disorders may include defective enzymes, defects in receptor and transport systems, alterations in the structure, function or quality of non-enzyme proteins, and genetically determined adverse reactions to drugs. Some of these conditions are related to familial hypercholesterolemia, lysosomal storage diseases, glycogen storage diseases and



neurofibromatosis. Provision of gene therapy using the method of the present invention may address, for example, supplementation of an animal with a protein or enzyme that the animal needs in view of its inadequate or faulty production of the protein. Practice of the present invention can be used to inactivate the defective gene  
5 through use of siRNA and then the transposon based vector can be used to insert the normal gene in order to restore function.

#### Other disorders

The present invention also provides gene therapy for animals that may not possess a demonstrable genetic deficiency. However, such animals may require  
10 supplementation of specific proteins that may be produced in inadequate amounts or in a defective form that renders them biologically inactive or marginally active. Alternatively, animals may produce too much of a protein that causes a disease or condition that renders the animal sick. Such animals may require gene therapy to reduce the transcription of a gene that makes the protein. Such animals may require  
15 gene therapy to produce proteins or peptides to blunt or block the activity of the overabundant protein.

#### Diseases and Conditions

Numerous diseases and conditions may be treated with the gene therapy  
20 method of the present invention, including, but not limited to, diseases and conditions of the following systems: cardiovascular system (atherosclerosis, hypercholesterolemia, disorders of LDL, HDL and apolipoprotein synthesis and metabolism, hypertension); reproductive system (reproductive health and dysfunction, fertility, infertility, menopause, menarche, puberty, superovulation, timing of  
25 ovulation, inducement of ovulation, inducement of sterilization (especially of companion animals), mastitis, cancers of the reproductive system); endocrine and neuroendocrine systems (hypopituitary disorders, hypothalamic disorders, hypogonadism, precocious puberty, dwarfism, infertility, lactation, diabetes, thyroid disease, adrenal cortical or adrenal medullary disease, appetite, feeding, drinking,  
30 temperature regulation); metabolic system (digestive disorders, inborn errors of metabolism, disorders of intermediate metabolism, fat metabolism, Crohn's disease; phenylketonuria, chronic wasting disease, phosphofructokinase deficiency, pyruvic kinase deficiency; nervous system (Parkinson's disease, Alzheimer's disease, Huntington's disease, encephalopathy, bovine spongiform encephalopathy, conditions

related to neurotransmitter transporter systems such as catecholamine transporters and reuptake mechanisms (serotonin, norepinephrine, dopamine) such as depression, psychosis, neurosis, addiction, alcoholism, motivation, bulimia, hyperphagia); immune system (feline immunodeficiency virus, simian immunodeficiency virus, immunodeficiency disorders including severe immunodeficiency disorders and severe combined immunodeficiency disorders, leukemia, autoimmune disorders, allergies, lupus, multiple sclerosis, scleroderma, disorders involving various immunoglobulins, interleukins, cytokines and lymphokines); hematologic and related disorders (sickle cell anemia, clotting disorders, von Willibrand's Disease); musculoskeletal system (arthritis, rheumatoid arthritis, osteoarthritis, muscular dystrophy); cancer (ovarian, prostate, breast, colon, brain, lung, kidney, skin); respiratory system (lung cancer, laryngeal cancer, cystic fibrosis); obesity; aging; cosmetic treatment of skin and hair; any form of cancer (skin (melanoma, basal, squamous), bladder, colon, stomach, esophageal, liver, pancreatic, testicular, prostate, ovarian, cervical, uterine, breast, lung, laryngeal, thyroid, adrenal, renal, penile, head, neck, brain (neural, glial); disorders involving receptors, particularly membrane bound receptors; and, infectious diseases (parasitic disease, bacterial infectious disease, viral disease, pneumovirus, Eastern equine encephalitis, West Nile virus, malaria, lyme disease, ehrlichiosis, retroviral infections, rabies, and diseases borne by invertebrates such as ticks, fleas, flies and mosquitoes.

The transposon-based vectors of the present invention can be used for the treatment of various genetic disorders. For example, one or more LTR-vector complexes can be administered to an animal for the treatment of a single gene disorder including, but not limited to, animal equivalents of Huntington's disease, alpha-1-antitrypsin deficiency Alzheimer's disease, various forms of breast cancer, cystic fibrosis, galactosemia, congenital hypothyroidism, maple syrup urine disease, neurofibromatosis 1, phenylketonuria, sickle cell disease, and Smith-Lemli-Opitz (SLO/RSH) Syndrome any metabolic errors, autoimmune diseases, shipping fever in cattle, mastitis, bacterial or viral diseases, alteration of skin pigment in animals, production of animals with enhanced growth characteristics and nutrient utilization. In these embodiments, the transposon-based vector contains a non-mutated, or non-disease causing form of the gene known to cause such disorder. The transposon-based vectors of the present invention can also be used to treat multiple gene disorders. The transposon-based vectors of the present invention can be used as DNA

vaccines and are useful in organ-specific disease treatments and localized disease treatments.

Preferably, the transposase contained within the transposase-based vector is operably linked to an inducible promoter such as a tissue specific promoter such that  
 5 the non-mutated gene of interest is inserted into a specific tissue wherein the mutated gene is expressed *in vivo*. Additionally, the DNA constructs of the present invention can be used to provide cells or tissues with "beacons", such as receptor molecules, for binding of therapeutic agents in order to provide tissue and cell specificity for the therapeutic agents. Several promoters and exogenous genes can be combined in one  
 10 vector to produce progressive, controlled, treatments, from a single vector delivery.

In avians, for example, one or more LTR-vector complexes are administered to an avian for the treatment of a viral or bacterial infection/disease including, but not limited to, Colibacillosis (Coliform infections), Mycoplasmosis (CRD, Air sac, Sinusitis), Fowl Cholera, Necrotic Enteritis, Ulcerative Enteritis (Quail disease),  
 15 Pullorum Disease, Fowl Typhoid, Botulism, Infectious Coryza, Erysipelas, Avian Pox, Newcastle Disease, Infectious Bronchitis, Quail Bronchitis, Lymphoid Leukosis, Marek's Disease (Visceral Leukosis), Infectious Bursal Disease (Gumboro), Avian Encephalomyelitis (AE, Avian Influenza (AI), Avian Leukosis Virus (LLAg, LLAb, ALV-J), Reticuloendotheliosis Virus (REV), Avian Pneumovirus (APV), Chicken  
 20 Anemia Virus (CAV), Infectious Bronchitis Virus (IBV), Infectious Bursal Disease Virus – Gumboro Disease (IBD, IBD-XR), Mycoplasma (MG, MS, MG/MS, MM), Newcastle Disease Virus (NDV, NDV-T), Ornithobacterium rhinotracheale (ORT), Pasteurella multocida (PM, PM-T), Reovirus (REO), and Salmonella enteritidis (SE).

In swine, for example, one or more transposon-based vectors are administered  
 25 for the treatment of a viral or bacterial infection/disease including, but not limited to, Pseudorabies Virus – Aujeszky's Disease (PRV-V, PRV-S, PRV gl (gE)), Porcine Reproductive and Respiratory Syndrome (PRRS 2XR), Classical Swine Fever Virus (CSFV Ab, CSFV Ag), Swine Influenza (SIV H1N1), Mycoplasma hyopneumoniae (M. hyo.), and Swine Salmonella.

30 In ruminants, for example, one or more transposon-based vectors are administered for the treatment of a viral or bacterial infection/disease including, but not limited to, Bovine Leukemia Virus (BLV), Infectious Bovine Rhinotracheitis (IBR, IBR gB, IBR gE), Brucella abortus (B. abortus), Mycobacterium

paratuberculosis – Johne's Disease (M. pt.), Neospora caninum, and Bovine Viral Diarrhea Virus (BVDV).

In horses, for example, one or more transposon-based vectors are administered for the treatment of a viral or bacterial infection/disease including, but not limited to,  
5 Equine Infectious Anemia (EIA).

Numerous genetic diseases that affect humans are shown in Table 3.

Methods of Administering Compositions Comprising Transposon-Based Vectors to Provide Therapy

10 The compositions of the present invention, comprising a vector, a transfecting reagent and an acceptable carrier may be delivered to a desired location in an animal receiving gene therapy through administration via a selected route. Accordingly, the compositions may be administered in a variety of ways including, but not limited to the following: through a vascular system, a duct system, within the lumen of an organ,  
15 into an organ, tissue or cell, into a body cavity, into the cerebrospinal fluid, topically, through the gastrointestinal system, through the reproductive system, through the urinary system, intraperitoneally, and through the respiratory system.

The vector can be administered into the vascular system. In a preferred embodiment, the vector is administered into the cardiovascular system and  
20 specifically into one or more chambers of the heart. Administration of the vector into the cardiovascular system and specifically into one or more chambers of the heart, results in the distribution of the vector to the organs and tissues and cells receiving blood supply from the vessel or the heart. In a preferred embodiment, administration of the vector into the left ventricle of the heart results in distribution of the vector to  
25 the organs supplied by branches of the aorta, for example the celiac, gonadal, superior (cranial) mesenteric and inferior (caudal) mesenteric arteries. Such distribution targets include the liver, ovary, oviduct and testes, among other organs. Administration through the internal mammary artery transfects secretory cells of the lactating mammary gland to perform a desired function, such as to synthesize and  
30 secrete a desired protein or peptide into the milk. Administration through the internal mammary artery would also target breast cancer cells. Administration of the compositions into the artery supplying the ovary or to the fallopian tube to supply those tissues. In this manner, follicles are transfected to create a germline transgenic animal. Alternatively, supplying the compositions through the artery leading to the

fallopian tube preferably transfects the epithelial cells. Such transfected epithelial cells manufacture a desired protein or peptide for deposition in the egg white. Administration of the compositions through the portal vein or hepatic artery targets uptake and transformation of hepatic cells. Intravascular administration further includes administration in to any vein, including but not limited to veins in the systemic circulation and veins in the hepatic portal circulation. Intravascular administration further includes administration into the cerebrovascular system, including the carotid arteries, the vertebral arteries and branches thereof.

Intravascular administration may be coupled with methods known to influence the permeability of vascular barriers such as the blood brain barrier and the blood testes barrier, in order to enhance transfection of cells that are difficult to affect through vascular administration. Such methods are known to one of ordinary skill in the art and include use of hyperosmotic agents, mannitol, hypothermia, nitric oxide, alkylglycerols, lipopolysaccharides (Haluska et al., Clin. J. Oncol. Nursing 8(3): 263-267, 2004; Brown et al., Brain Res., 1014: 221-227, 2004; Ikeda et al., Acta Neurochir. Suppl. 86:559-563, 2004; Weyerbrock et al., J. Neurosurg. 99(4):728-737, 2003; Erdlenbruch et al., Br. J. Pharmacol. 139(4):685-694, 2003; Gaillard et al., Microvasc. Res. 65(1):24-31, 2003; Lee et al., Biol. Reprod. 70(2):267-276, 2004)).

Intravascular administration may also be coupled with methods known to influence vascular diameter, such as use of beta blockers, nitric oxide generators, prostaglandins and other reagents that increase vascular diameter and blood flow.

In one embodiment, the animal is an egg-laying animal, and more preferably, an avian, and the transposon-based vectors comprising the polynucleotide cassettes are administered into the vascular system, preferably into the heart. In one embodiment, between approximately 1 and 300  $\mu\text{g}$ , 1 and 200  $\mu\text{g}$ , 5 and 200  $\mu\text{g}$ , or 5 and 150  $\mu\text{g}$  of a transposon-based vector containing the polynucleotide cassette is administered to the vascular system, preferably into the heart. In a chicken, it is preferred that between approximately 1 and 300  $\mu\text{g}$ , or 5 and 200  $\mu\text{g}$  are administered to the vascular system, preferably into the heart, more preferably into the left ventricle. The total injection volume for administration into the left ventricle of a chicken may range from about 10  $\mu\text{l}$  to about 3.0 ml, or from about 100  $\mu\text{l}$  to about 1.5 ml, or from about 200  $\mu\text{l}$  to about 1.0 ml, or from about 200  $\mu\text{l}$  to about 800  $\mu\text{l}$ . It is to be understood that the total injection volume may vary depending on the duration

of the injection. Longer injection durations may accommodate higher total volumes. In a quail, it is preferred that between approximately 1 and 200  $\mu\text{g}$ , or 5 and 150  $\mu\text{g}$  are administered to the vascular system, preferably into the heart, more preferably into the left ventricle. The total injection volume for administration into the left ventricle of a quail may range from about 10  $\mu\text{l}$  to about 1.0 ml, or from about 100  $\mu\text{l}$  to about 800  $\mu\text{l}$ , or from about 200  $\mu\text{l}$  to about 600  $\mu\text{l}$ . It is to be understood that the total injection volume may vary depending on the duration of the injection. Longer injection durations may accommodate higher total volumes. The microgram quantities represent the total amount of the vector with the transfection reagent.

Other, non-avian animals will require different volumes and amounts for injection and these values can be extrapolated on a body weight or surface area basis as known to one of ordinary skill in the art. For example, an intravascular administration into a rat may occur through a cannula inserted into the right or left atrium or ventricle and may comprise a volume of from about 0.05 ml to 4 ml containing 1 and 300  $\mu\text{g}$  is injected gradually.

Administration may also occur through non vascular routes. For example, administration through the urethra and into the bladder targets the transitional epithelium of the bladder. Administration through the vagina and cervix targets the lining of the uterus. For example, administration may occur directly into a muscle to transfect striated muscle cells for production of a desired protein.

In one embodiment of the present invention, a transposon-based vector comprising a gene encoding proinsulin is administered to diabetic animals receiving gene therapy for incorporation into liver cells in order to treat or cure diabetes. The specific incorporation of the proinsulin gene into the liver is accomplished by placing the transposase of the transposon-based vector under control of liver-specific promoter, such as the glucose-6-phosphatase promoter (G6P). This approach is useful for treatment of both type I and type II diabetes. The G6P promoter has been shown to be glucose responsive (Arguad, D., et al. 1996, *Diabetes* 45: 1563-1571), and thus, glucose-regulated insulin production is achieved using DNA constructs of the present invention. Integrating a proinsulin gene into liver cells circumvents the problem of destruction of pancreatic islet cells in the course of type 1 diabetes.

In another embodiment, shortly after diagnosis of type I diabetes, the cells of the immune system destroying  $\beta$ -cells of the pancreas are selectively removed using

the DNA constructs of the present invention, thus allowing normal  $\beta$ -cells to repopulate the pancreas.

For treatment of type II diabetes, the DNA constructs of the present invention are specifically incorporated into the pancreas by placing the transposase of the transposon-based vector under the control of a pancreas-specific promoter, such as an insulin promoter. In this embodiment, the vector is delivered to a diabetic animal via injection into an artery supplying the pancreas. For delivery, the vector is complexed with a transfection agent. The artery distributes the complex throughout the pancreas, where individual cells receive the vector DNA. Following uptake into the target cell, the insulin promoter is recognized by transcriptional machinery of the cell, the transposase encoded by the vector is expressed, and stable integration of the proinsulin gene occurs. It is expected that a small percentage of the DNA construct would be transported to other tissues, and that these tissues would be transfected. However, these tissues would not be stably transfected due to failure of these other cells to activate the insulin promoter. The DNA would likely be lost when the cell dies or degraded over time.

In addition to the transposon-based vectors described above, the present invention also includes methods of administering the transposon-based vectors to an animal, methods of producing a transgenic animal wherein a gene of interest is incorporated into the germline of the animal and methods of producing a transgenic animal wherein a gene of interest is incorporated into cells other than the germline cells (somatic cells) of the animal. For example, the transposon-based vectors of the present invention are administered to a reproductive organ of an animal via any method known to those of skill in the art. Preferred reproductive organs include a testis, an ovary, an oviduct, a mammary gland, and a fallopian tube.

In some embodiments, a transposon-based vector is directly administered to the reproductive organ. Direct administration encompasses injection into the organ, and in one embodiment, a transposon-based vector is injected into the lumen of the oviduct, and more preferably, the lumen of the magnum or the infundibulum of the oviduct. The transposon-based vectors may additionally or alternatively be placed in an artery supplying the reproductive organ. Administering the vectors to the artery supplying the ovary results in transfection of follicles and oocytes in the ovary to create a germline transgenic animal. Alternatively, supplying the vectors through an

artery leading to the oviduct would preferably transfect the tubular gland and epithelial cells. Such transfected cells manufacture a desired protein or peptide for deposition in the egg white. In one embodiment, a transposon-based vector is administered into the lumen of the magnum or the infundibulum of the oviduct and to  
5 an artery supplying the oviduct. Indirect administration to the oviduct epithelium may occur through the cloaca. Direct administration into the mammary gland may be achieved through introduction into the duct system of the mammary gland or an artery supplying the mammary gland.

The transposon-based vectors may be administered in a single administration,  
10 multiple administrations, continuously, or intermittently. The transposon-based vectors may be administered by injection, via a catheter, an osmotic mini-pump or any other method. In some embodiments, the transposon-based vector is administered to an animal in multiple administrations, each administration containing the vector and a different transfecting reagent.

15 The transposon-based vectors may be administered to the animal at any desirable time for gene therapy during the lifetime of the animal.

In one embodiment, between approximately 1  $\mu$ g and 5 mg, 1  $\mu$ g and 3 mg, 1  $\mu$ g and 1 mg, of transposon-based vector DNA is administered to the animal. Intraoviduct administration of the transposon-based vectors of the present invention  
20 resulted in incorporation of the gene of interest into the cells of the oviduct as evidenced by a PCR positive signal in the oviduct tissue, demonstrating that the present invention is effective in providing genetic therapy to the animal. In other embodiments, the transposon-based vector is administered to an artery that supplies the oviduct. These methods of administration may also be combined with any  
25 methods for facilitating transfection, including without limitation, electroporation, gene guns, injection of naked DNA, and use of dimethyl sulfoxide (DMSO).

According to the present invention, the transposon-based vector is administered in conjunction with an acceptable carrier and/or transfection reagent. Acceptable carriers include, but are not limited to, water, saline, Hanks Balanced Salt  
30 Solution (HBSS), Tris-EDTA (TE) and lyotropic liquid crystals. Transfection reagents commonly known to one of ordinary skill in the art that may be employed include, but are not limited to, the following: cationic lipid transfection reagents, cationic lipid mixtures, polyamine reagents, liposomes and combinations thereof;



SUPERFECT®, Cytofectene, BioPORTER®, GenePORTER®, NeuroPORTER®, and perfectin from Gene Therapy Systems; lipofectamine, cellfectin, DMRIE-C oligofectamine, and PLUS reagent from InVitrogen; Xtreme gene, fugene, DOSPER and DOTAP from Roche; Lipotaxi and Genejammer from Strategene; and Escort  
5 from SIGMA. In one embodiment, the transfection reagent is SUPERFECT®. The ratio of DNA to transfection reagent may vary based upon the method of administration. In one embodiment, the transposon-based vector is administered to the oviduct and the ratio of DNA to transfection reagent can be from 1:1.5 to 1:15, preferably 1:2 to 1:5, all expressed as wt/vol. Transfection may also be accomplished  
10 using other means known to one of ordinary skill in the art, including without limitation electroporation, gene guns, injection of naked DNA, and use of dimethyl sulfoxide (DMSO).

Depending upon the cell or tissue type targeted for transfection, the form of the transposon-based vector may be important. Plasmids harvested from bacteria are  
15 generally closed circular supercoiled molecules, and this is the preferred state of a vector for gene delivery because of the ease of preparation. In some instances, transposase expression and insertion may be more efficient in a relaxed, closed circular configuration or in a linear configuration. In still other instances, a purified transposase protein may be co-injected with a transposon-based vector containing the  
20 gene of interest for more immediate insertion. This could be accomplished by using a transfection reagent complexed with both the purified transposase protein and the transposon-based vector.

#### Testing for and Breeding Animals Carrying the Transgene

25 Following administration of a transposon-based vector to an animal receiving gene therapy, DNA is extracted from the animal to confirm integration of the gene of interest. Advantages provided by the present invention include the high rates of integration, or incorporation, and transcription of the gene of interest when administered to a bird via an intraoviduct or intraovarian route (including intraarterial  
30 administrations to arteries leading to the oviduct or ovary). The construct of Figure 2, when administered to Japanese quail hens, resulted in expression of the fusion peptide in the oviduct cells and subsequent secretion and deposition in the egg white. Assaying of the egg white on and SDS PAGE gel demonstrated the presence of the expressed protein. The sequence of the fusion protein was verified by MALDI-TOF

analysis by an independent third party. The proinsulin/ENT TAG protein from a transgenic hen was isolated following ammonium sulfate precipitation and ion exchange chromatography. The transposon-based vector was successfully administered to a hen, and the gene of interest successfully integrated. The protein  
5 encoded by the gene of interest was produced and deposited in egg white produced by the transgenic hen.

Actual frequencies of integration may be estimated both by comparative strength of the PCR signal, and by histological evaluation of the tissues by quantitative PCR. Another method for estimating the rate of transgene insertion is the  
10 so-called primed *in situ* hybridization technique (PRINS). This method determines not only which cells carry a transgene of interest, but also into which chromosome the gene has inserted, and even what portion of the chromosome. Briefly, labeled primers are annealed to chromosome spreads (affixed to glass slides) through one round of PCR, and the slides are then developed through normal *in situ* hybridization  
15 procedures. This technique combines the best features of *in situ* PCR and fluorescence *in situ* hybridization (FISH) to provide distinct chromosome location and copy number of the gene in question. The 28s rRNA gene will be used as a positive control for spermatogonia to confirm that the technique is functioning properly. Using different fluorescent labels for the transgene and the 28s gene causes cells  
20 containing a transgene to fluoresce with two different colored tags.

Breeding experiments are also conducted to determine if germline transmission of the transgene has occurred. In a general bird breeding experiment performed according to the present invention, each male bird was exposed to 2-3 different adult female birds for 3-4 days each. This procedure was continued with  
25 different females for a total period of 6-12 weeks. Eggs are collected daily for up to 14 days after the last exposure to the transgenic male, and each egg is incubated in a standard incubator. The resulting embryos are examined for transgene presence at day 3 or 4 using PCR. It is to be understood that the above procedure can be modified to suit animals other than birds and that selective breeding techniques may be  
30 performed to amplify gene copy numbers and protein output.

#### Production of Desired Proteins or Peptides in Egg White

In one embodiment, the transposon-based vectors of the present invention may be administered to a bird receiving gene therapy for production of desired proteins or

peptides in the egg white. These transposon-based vectors preferably contain one or more of an ovalbumin promoter, an ovomucoid promoter, an ovalbumin signal sequence and an ovomucoid signal sequence. Oviduct-specific ovalbumin promoters are described in B. O'Malley et al., 1987. EMBO J., vol. 6, pp. 2305-12; A. Qiu et al.,  
5 1994. Proc. Nat. Acad. Sci. (USA), vol. 91, pp. 4451-4455; D. Monroe et al., 2000. Biochim. Biophys. Acta, 1517 (1):27-32; H. Park et al., 2000. Biochem., 39:8537-8545; and T. Muramatsu et al., 1996. Poult. Avian Biol. Rev., 6:107-123. Examples of transposon-based vectors designed for production of a desired protein in an egg white are shown in Figures 2 and 3.

10

#### Production of Desired Proteins or Peptides in Egg Yolk

The present invention is particularly advantageous for production of recombinant peptides and proteins of low solubility in the egg yolk. Such proteins include, but are not limited to, membrane-associated or membrane-bound proteins,  
15 lipophilic compounds; attachment factors, receptors, and components of second messenger transduction machinery. Low solubility peptides and proteins are particularly challenging to produce using conventional recombinant protein production techniques (cell and tissue cultures) because they aggregate in water-based, hydrophilic environments. Such aggregation necessitates denaturation and re-  
20 folding of the recombinantly-produced proteins, which may deleteriously affect their structure and function. Moreover, even highly soluble recombinant peptides and proteins may precipitate and require denaturation and renaturation when produced in sufficiently high amounts in recombinant protein production systems. The present invention provides an advantageous resolution of the problem of protein and peptide  
25 solubility during production of large amounts of recombinant proteins.

In one embodiment of the present invention wherein germline transfection is obtained via intraovarian administration of the transposon-based vector, deposition of a desired protein into the egg yolk is accomplished in offspring by attaching a sequence encoding a protein capable of binding to the yolk vitellogenin receptor to a  
30 gene of interest that encodes a desired protein. This transposon-based vector can be used for the receptor-mediated uptake of the desired protein by the oocytes. In a preferred embodiment, the sequence ensuring the binding to the vitellogenin receptor is a targeting sequence of a vitellogenin protein. The invention encompasses various vitellogenin proteins and their targeting sequences. In a preferred embodiment, a

chicken vitellogenin protein targeting sequence is used, however, due to the high degree of conservation among vitellogenin protein sequences and known cross-species reactivity of vitellogenin targeting sequences with their egg-yolk receptors, other vitellogenin targeting sequences can be substituted. One example of a construct for use in the transposon-based vectors of the present invention and for deposition of an insulin protein in an egg yolk is a transposon-based vector containing a vitellogenin promoter, a vitellogenin targeting sequence, a TAG sequence, a pro-insulin sequence and a synthetic polyA sequence. The present invention includes, but is not limited to, vitellogenin targeting sequences residing in the N-terminal domain of vitellogenin, particularly in lipovitellin I. In one embodiment, the vitellogenin targeting sequence contains the polynucleotide sequence of SEQ ID NO:28. In a preferred embodiment, the transposon-based vector contains a transposase gene operably-linked to a constitutive promoter and a gene of interest operably-linked to a liver-specific promoter and a vitellogenin targeting sequence.

The following examples will serve to further illustrate the present invention without, at the same time, however, constituting any limitation thereof. On the contrary, it is to be clearly understood that resort may be had to various embodiments, modifications and equivalents thereof which, after reading the description herein, may suggest themselves to those skilled in the art without departing from the spirit of the invention.

#### EXAMPLE 1

##### *Preparation of Transposon-Based Vector pTnMod*

A vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells, given below as SEQ ID NO:7. The vector of SEQ ID NO:7, termed pTnMod, was constructed and its sequence verified.

This vector employed a cytomegalovirus (CMV) promoter. A modified Kozak sequence (ACCATG) (SEQ ID NO:8) was added to the promoter. The nucleotide in the wobble position in nucleotide triplet codons encoding the first 10 amino acids of transposase was changed to an adenine (A) or thymine (T), which did not alter the amino acid encoded by this codon. Two stop codons were added and a synthetic polyA was used to provide a strong termination sequence. This vector uses a promoter designed to be active soon after entering the cell (without any induction) to increase the likelihood of stable integration. The additional stop codons and synthetic

polyA insures proper termination without read through to potential genes downstream.

The first step in constructing this vector was to modify the transposase to have the desired changes. Modifications to the transposase were accomplished with the primers High Efficiency forward primer (Hef) Altered transposase (ATS)-Hef 5' ATCTCGAGACCATGTGTGAACTTGATATTTTACATGATTCTCTTTACC 3' (SEQ ID NO:42) and Altered transposase- High efficiency reverse primer (Her) 5' GATTGATCATTATCATAATTTCCCCAAAGCGTAACC 3' (SEQ ID NO:43, a reverse complement primer). The sequence ACCATG (SEQ ID NO:8) contains the Kozak sequence and start codon for the transposase and the underlined bases represent changes in the wobble position to an A or T of codons for the first 10 amino acids (without changing the amino acid coded by the codon). Primer ATS-Her (SEQ ID NO:43) contains an additional stop codon TAA in addition to native stop codon TGA and adds a Bcl I restriction site to allow directional cloning. These primers were used in a PCR reaction with pTnLac (p defines plasmid, tn defines transposon, and lac defines the beta fragment of the lactose gene, which contains a multiple cloning site) as the template for the transposase and a FailSafe™ PCR System (which includes enzyme, buffers, dNTP's, MgCl<sub>2</sub> and PCR Enhancer; Epicentre Technologies, Madison, WI). Amplified PCR product was electrophoresed on a 1% agarose gel, stained with ethidium bromide, and visualized on an ultraviolet transilluminator. A band corresponding to the expected size was excised from the gel and purified from the agarose using a Zymo Clean Gel Recovery Kit (Zymo Research, Orange, CA). Purified DNA was digested with restriction enzymes Xho I (5') and Bcl I (3') (New England Biolabs, Beverly, MA) according to the manufacturer's protocol. Digested DNA was purified from restriction enzymes using a Zymo DNA Clean and Concentrator kit (Zymo Research).

Plasmid gWhiz (Gene Therapy Systems, San Diego, CA) was digested with restriction enzymes Sal I and BamH I (New England Biolabs), which are compatible with Xho I and Bcl I, but destroy the restriction sites. Digested gWhiz was separated on an agarose gel, the desired band excised and purified as described above. Cutting the vector in this manner facilitated directional cloning of the modified transposase (mATS) between the CMV promoter and synthetic polyA.

To insert the mATS between the CMV promoter and synthetic polyA in gWhiz, a Stratagene T4 Ligase Kit (Stratagene, Inc. La Jolla, CA) was used and the

ligation set up according to the manufacturer's protocol. Ligated product was transformed into *E. coli* Top10 competent cells (Invitrogen Life Technologies, Carlsbad, CA) using chemical transformation according to Invitrogen's protocol. Transformed bacteria were incubated in 1 ml of SOC (GIBCO BRL, CAT# 15544-042) medium for 1 hour at 37° C before being spread to LB (Luria-Bertani media (broth or agar)) plates supplemented with 100 µg/ml ampicillin (LB/amp plates). These plates were incubated overnight at 37° C and resulting colonies picked to LB/amp broth for overnight growth at 37° C. Plasmid DNA was isolated using a modified alkaline lysis protocol (Sambrook et al., 1989), electrophoresed on a 1% agarose gel, and visualized on a U.V. transilluminator after ethidium bromide staining. Colonies producing a plasmid of the expected size (approximately 6.4 kbp) were cultured in at least 250 ml of LB/amp broth and plasmid DNA harvested using a Qiagen Maxi-Prep Kit (column purification) according to the manufacturer's protocol (Qiagen, Inc., Chatsworth, CA). Column purified DNA was used as template for sequencing to verify the changes made in the transposase were the desired changes and no further changes or mutations occurred due to PCR amplification. For sequencing, Perkin-Elmer's Big Dye Sequencing Kit was used. All samples were sent to the Gene Probes and Expression Laboratory (LSU School of Veterinary Medicine) for sequencing on a Perkin-Elmer Model 377 Automated Sequencer.

Once a clone was identified that contained the desired mATS in the correct orientation, primers CMVf-NgoM IV and Syn-polyA-BstE II were used to PCR amplify the entire CMV promoter, mATS, and synthetic polyA for cloning upstream of the transposon in pTnLac. The PCR was conducted with FailSafe™ as described above, purified using the Zymo Clean and Concentrator kit, the ends digested with NgoM IV and BstE II (New England Biolabs), purified with the Zymo kit again and cloned upstream of the transposon in pTnLac as described below.

Plasmid pTnLac was digested with NgoM IV and BstE II to remove the ptac promoter and transposase and the fragments separated on an agarose gel. The band corresponding to the vector and transposon was excised, purified from the agarose, and dephosphorylated with calf intestinal alkaline phosphatase (New England Biolabs) to prevent self-annealing. The enzyme was removed from the vector using a Zymo DNA Clean and Concentrator-5. The purified vector and CMVp/mATS/polyA

were ligated together using a Stratagene T4 Ligase Kit and transformed into *E. coli* as described above.

Colonies resulting from this transformation were screened (mini-preps) as describe above and clones that were the correct size were verified by DNA sequence  
5 analysis as described above. The vector was given the name pTnMod (SEQ ID NO:7) and includes the following components:

Base pairs 1-130 are a remainder of F1(-) on from pBluescriptII sk(-) (Stratagene), corresponding to base pairs 1-130 of pBluescriptII sk(-).

Base pairs 131 - 132 are a residue from ligation of restriction enzyme sites  
10 used in constructing the vector.

Base pairs 133 -1777 are the CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems), corresponding to bp 229-1873 of pGWiz. The CMV promoter was modified by the addition of an ACC sequence upstream of ATG.

Base pairs 1778-1779 are a residue from ligation of restriction enzyme sites  
15 used in constructing the vector.

Base pairs 1780 - 2987 are the coding sequence for the transposase, modified from Tn10 (GenBank accession J01829) by optimizing codons for stability of the transposase mRNA and for the expression of protein. More specifically, in each of the codons for the first ten amino acids of the transposase, G or C was changed to A or T  
20 when such a substitution would not alter the amino acid that was encoded.

Base pairs 2988-2993 are two engineered stop codons.

Base pair 2994 is a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 2995 - 3410 are a synthetic polyA sequence taken from the pGWiz  
25 vector (Gene Therapy Systems), corresponding to bp 1922-2337 of 10 pGWiz.

Base pairs 3415 - 3718 are non-coding DNA that is residual from vector pNK2859.

Base pairs 3719 - 3761 are non-coding  $\lambda$  DNA that is residual from pNK2859.

Base pairs 3762 - 3831 are the 70 bp of the left insertion sequence recognized  
30 by the transposon Tn10.

Base pairs 3832-3837 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 3838 - 4527 are the multiple cloning site from pBluescriptII sk(20), corresponding to bp 924-235 of pBluescriptII sk(-). This multiple cloning site may be used to insert any coding sequence of interest into the vector.

Base pairs 4528-4532 are a residue from ligation of restriction enzyme sites  
5 used in constructing the vector.

Base pairs 4533 - 4602 are the 70 bp of the right insertion sequence recognized by the transposon Tn10.

Base pairs 4603 - 4644 are non-coding  $\lambda$  DNA that is residual from pNK2859.

Base pairs 4645 - 5488 are non-coding DNA that is residual from pNK2859.

10 Base pairs 5489 - 7689 are from the pBluescriptII sk(-) base vector - (Stratagene, Inc.), corresponding to bp 761-2961 of pBluescriptII sk(-).

Completing pTnMod is a pBlueScript backbone that contains a *colE1* origin of replication and an antibiotic resistance marker (ampicillin).

It should be noted that all non-coding DNA sequences described above can be  
15 replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

All plasmid DNA was isolated by standard procedures. Briefly, *Escherichia coli* containing the plasmid was grown in 500 mL aliquots of LB broth (supplemented with an appropriate antibiotic) at 37°C overnight with shaking. Plasmid DNA was  
20 recovered from the bacteria using a Qiagen Maxi-Prep kit (Qiagen, Inc., Chatsworth, CA) according to the manufacturer's protocol. Plasmid DNA was resuspended in 500  $\mu$ L of PCR-grade water and stored at -20°C until used.

## EXAMPLE 2

### 25 *Transposon-Based Vector pTnMCS*

Another transposon-based vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells. This vector was termed pTnMCS and its constituents are provided below. The sequence of the pTnMCS vector is provided in SEQ ID NO:6. The pTnMCS vector contains an avian optimized polyA sequence  
30 operably-linked to the transposase gene. The avian optimized polyA sequence contains approximately 40 nucleotides that precede the A nucleotide string.

Bp 1 - 130 Remainder of F1 (-) ori of pBluescriptII sk(-) (Stratagene) bp1-130

Bp 133 - 1777 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy Systems) bp 229-1873



- Bp 1783 – 2991 Transposase, from Tn10 (GenBank accession #J01829) bp 108-1316  
 Bp 2992 – 3344 Non coding DNA from vector pNK2859  
 Bp 3345 – 3387 Lambda DNA from pNK2859  
 Bp 3388 – 3457 70 bp of IS10 left from Tn10
- 5 Bp 3464 – 3670 Multiple cloning site from pBluescriptII sk(-), thru the XmaI site bp 924-718
- Bp 3671 - 3715 Multiple cloning site from pBluescriptII sk(-), from the XmaI site thru the XhoI site. These base pairs are usually lost when cloning into pTnMCS bp 717-673
- 10 Bp 3716 – 4153 Multiple cloning site from pBluescriptII sk(-), from the XhoI site bp 672-235
- Bp 4159 - 4228 70 bp of IS10 right from Tn10  
 Bp 4229 - 4270 Lambda DNA from pNK2859  
 Bp 4271 – 5114 Non-coding DNA from pNK2859
- 15 Bp 5115 - 7315 pBluescript sk (-) base-vector (Stratagene, Inc.) bp 761-2961.

### EXAMPLE 3

#### *Gene Therapy to Treat Cancer in an Animal*

#### Preparation of the transposon-based vector

- 20 The following genes were cloned into the transposon-based vector of the present invention: an SV40 promoter linked to the preprosequence of cecropin B together with either a gene of interest encoding Phor14:beta human chorionic gonadotropin (bHCG), a gene of interest encoding gonadotropin releasing hormone (GnRH):Phor 11, or a gene of interest encoding GnRH:Phor 14, each linked to a
- 25 cecropin B poly A. The Phor peptides are lytic peptides. In this manner, three different vectors were created: 1) pTnPhor14:bHCG; 2) pTnGnRH:Phor 11; and, 3) pTnGnRH:Phor 14. The base vector used was pBTnLac. The SV40 promoter is a constitutive promoter expressed at a moderate level. The cecropin B prepro peptide was selected to permit a peptide to be transported out of the cell. The cecropin B poly
- 30 A was selected to terminate mRNA synthesis. These vectors in turn stimulate production of the fusion peptides, GnRH:Phor 11 (SEQ ID NO:44), GnRH:Phor 14 (SEQ ID NO:45), and Phor14:bHCG (SEQ ID NO:46). These transposon-based vectors were designed to provide an alternative to conventional chemotherapy in animals with tumors that express a receptor for luteinizing hormone (LH) or GnRH at

their surface, for example prostatic, ovarian, breast, pancreatic and some small cell lung carcinomas.

Mice, fish, cats and dogs have received these compositions without any adverse side effects. Temporary sterility was induced in these animals as evidenced  
5 by a disrupted reproductive cycle.

The goal of this gene therapy was to administer the transposon-based vector complexed with a transfecting agent to the animal through any desired route (for example intravenous, intraperitoneal, intraarterial, or intramuscular) so that the animal makes the fusion protein. Next, the cells of the animal that expressed a receptor for  
10 LH or GnRH recognized and bound the LH or GnRH component of the fusion peptide and delivered the fusion peptide containing the lytic peptide component to the cell, eventually resulting in lysis of the cell.

This approach to gene therapy permits very specific targeting of cells for destruction since cells expressing receptors specific for a ligand are affected without  
15 affecting other cells as in conventional chemotherapy or radiotherapy. Further, this approach permits sustained delivery of the fusion peptides over time since the transgene is stably incorporated.

Use of the transposon-based vector in gene therapy for treatment of cancer in a dog  
20 and in a cat

A dog with breast cancer metastasized throughout the body was treated. An aged female retriever of about 85 pounds body weight was diagnosed with widespread inflammatory mammary carcinoma. The initial diagnosis was confirmed with a skin biopsy of nodular lesions on the left lateral chest wall. The biopsy demonstrated  
25 tumor nodules surrounded by fibrous tissue and tumor emboli within dermal lymphatic vessels. The tumor was composed of large, irregular, darkly blue stained cells with nuclear atypia and a high mitotic rate. The dog was administered the genetic constructs encoding for SEQ ID NO: 45 and SEQ ID NO: 46, i.v., at a dose of 50 ug of each construct in about 150 ul of transfection reagent (Superfect). Ten  
30 days later the dog suddenly died. Within three hours, the skin lesion near the original biopsy site was removed and fixed. Histological analysis of sections from this biopsy revealed advanced and severe necrosis of the tumor cells both within the fibrous lesions and within the lymphatic vessels. Adjacent, normal non-neoplastic cells and tissues were non-necrotic. The death of tumor cells was estimated at more than 90%.

A cat with breast cancer metastasized throughout the body was treated. The cat was diagnosed with widespread inflammatory mammary carcinoma. The initial diagnosis was confirmed with a skin biopsy of nodular lesions. The cat was administered the genetic constructs encoding for SEQ ID NO: 45 and SEQ ID NO: 46, i.v. at a dose of 50 ug of each construct in about 150 ul of transfection reagent (Superfect). A subsequent histological analysis of sections from this biopsy revealed advanced and severe necrosis of the tumor cells both within the fibrous lesions and within the lymphatic vessels. Adjacent, normal non-neoplastic cells and tissues were non-necrotic.

The results demonstrate the efficiency of the gene therapy method of the present invention to treat cancer in an animal.

#### EXAMPLE 4

*Development of a vector for tissue-specific insulin gene incorporation into animal liver.*

Figure 6 shows a scheme of a pTnMod-based vector for targeting an insulin gene into the liver. Using transposase (ATS) under control of liver-specific promoter, such as liver glucose-6-phosphatase (G6P) promoter, allows for tissue-specific incorporation of the insulin gene in the liver. The insulin gene is also placed under control of a glucose-6 phosphatase (G6P) promoter.

The G6P promoter is cloned from rat genomic liver DNA. Rat genomic liver DNA is prepared according to procedures known to one of ordinary skill in the art. The promoter is cloned by amplifying the gene sequence using specific primers in a PCR reaction using methods known to one of ordinary skill in the art.

Alternatively, rat G6P promoter sequence is deduced from rat G6P gene untranslated upstream region provided in GenBank accession number U57552.1 and a corresponding synthetic oligonucleotide is prepared by methods known to one of ordinary skill in the art, preferably by any one of a number of commercial suppliers of synthetic oligonucleotides.

The gene encoding human proinsulin is amplified from human cDNA according to methods known in the art, for example, by using PCR with the primers specific for a proinsulin gene sequence, which is shown in SEQ ID NO:29. Briefly, total mRNA is isolated from a human pancreas according to procedures standard in the art. This mRNA is used to produce cDNA according to procedures standard in the

art. Alternatively, the promoter sequence is amplified by PCR from a commercially available human pancreatic cDNA library, such as the one supplied by Clontech Laboratories, Inc. under 2000 catalog number 7115-1, Human Pancreas QUICK-Clone cDNA.

5 Proinsulin is PCR-amplified from cDNA using specific primers designed in accordance with the proinsulin DNA sequence. The gene encoding proinsulin is cloned into the multiple cloning site (MCS) of pGWiz downstream of the G6P promoter sequence and upstream of the polyadenylation sequence (polyA).

Each of the identified components is sequenced, and cloned into pTnMod  
10 according to the scheme shown in Figure 6. Transposase (ATS) of the pTnMod vector is also placed under the control of the G6P promoter, which is obtained as described above, by subcloning G6P promoter sequence upstream of the pTnMod transposase sequence. Insertion sequences are denoted IS.

Any other desired components are prepared and incorporated into the vector  
15 by methods known to one of ordinary skill in the art. This vector is termed pTnModIns. Sufficient amounts of substantially pure TnModIns DNA are prepared using methods common in the art.

#### EXAMPLE 5

20 *Treatment of rats with the vector for tissue-specific insulin gene incorporation*

Diabetic rats are obtained made diabetic by administering the drug streptozotocin (Zanosar; Upjohn, Kalamazoo, MI) at approximately 200 mg/kg.

The rats are bred and maintained according to standard procedures. pTnModIns DNA, an appropriate carrier, and, optionally, a transfection agent, are  
25 injected into rats' singhepatic (if using G6P) artery with the purpose of stable transformation. Incorporation of the insulin gene into the rat genome and levels of insulin expression are ascertained by a variety of methods known in the art. Blood and tissue samples from live or sacrificed animals are tested. A combination of PCR, Southern and Northern blots, in-situ hybridization and related nucleic acid analysis  
30 methods are used to determine incorporation of the vector-derived proinsulin DNA and levels of transcription of the corresponding mRNA in various organs and tissues of the rats. A combination of SDS-PAGE gels, Western Blot analysis, radioimmunoassay, and ELISA and other methods known to one of ordinary skill in the art are used to determine the presence of insulin and the amount produced.

Additional transfections of pTnModIns are used to increase protein expression if the initial amounts of the expressed insulin are not satisfactory, or if the level of expression tapers off. The physiological condition of the rats is closely examined post-transfection to register positive or any negative effects of the gene therapy.

- 5 Animals are examined over extended periods of time post-transfection in order to monitor the stability of gene incorporation and protein expression.

#### EXAMPLE 6

##### 10 *Intracardiac Injection of a Transposon-based Vector for Gene Therapy and Production of Transgenic Quail*

Direct cardiac injection coupled with a transposon-based vector was used to provide direct incorporation into either liver, oviduct, or ovaries and progenitor cells of each. The technique may also be used to transform the progenitor cells (spermatogonia) in the testes to give rise to transgenic sperm. Stable incorporation of  
15 the vector DNA in progenitor cells results in long term production of transgenic liver cells, ova and oviduct cells, including tubular gland cells, and sperm; presumably for the life of the bird.

Five Japanese quail from Louisiana State University (LSU) stock and five from Bull Run stock were anesthetized and injected in the left ventricle of the heart in  
20 with 20 µg of the transposon-based vector and transfection reagent in a total volume of 0.35 ml. A needle approximately 5/8 inches (25 gauge) in length was used for injections of LSU quail. A needle approximately 1 inch (22 gauge) in length was used for injections of Bull Run quail. The needles were connected to a 1ml tuberculin syringe containing the transfection mixture. Birds were held in the hand with the keel  
25 up. Feathers in the area of the left breast were grasped and a few down feathers removed over the injection site. The area sprayed with ethanol.

The injector placed his left hand over the bird with the tip of the forefinger placed on the anterior tip of the keel. The thumb was used to palpate the triangle-shaped, posterior end of the caudolateral process of the sternum. The caudolateral  
30 process was followed forward to where it joined the body of the sternum. This marked the U shaped bony border of the lateral notch. The U of the lateral notch is formed by the thoracic process and the caudolateral process. While maintaining the thumb in the lateral notch, an imaginary line was drawn straight down from the forefinger. Another imaginary line was drawn at the angle of the caudolateral process

from the tip of the thumb forward. The site where the needle was placed was the intersection of these lines. This is approximately 2cm towards the bird's head from the tip of the thumb.

The needle and syringe were held parallel to the table. The needle was  
5 inserted into the superficial pectoralis muscle. Without completely withdrawing the needle, it was repositioned slightly to one side or the other until an intercostal space was found.

The needle was placed into the left breast muscle at about a 45° angle. When the needle was about halfway in, the needle hit the sternum. Next, the needle was  
10 partially removed and repositioned at a steeper angle until the sternum was no longer encountered. At this angle the needle dropped into the left ventricle of the heart. A flash of blood appeared in the syringe and pulsed in the hub at the rate of the heartbeat. The plunger on the syringe was slowly depressed. If there was an air bubble above the solution inside the syringe, the plunger was stopped before the air  
15 was pushed out into the blood. The needle was removed and disposed in a biohazard sharps container. The bird was returned to its cage and monitored for any signs of distress. (See A Color Atlas of Avian Anatomy. John McLelland. W.B. Saunders Company, 1991 for view of the anatomy of this area).

The vector CMVp/pp/HC/ProLys/LC/CPA (SEQ ID NO: 47), which encodes  
20 monoclonal antibody RM-2, was injected into the left ventricle of female Japanese quail. These birds were held for 2 days post-injection and sacrificed by cervical dislocation. Immediately after sacrifice, the visceral cavity of each bird was opened and a piece of liver, ovary, and oviduct was removed. For oviduct, a section from the magnum was removed and scissors were used to make a longitudinal cut that opened  
25 the tube and allowed it to lay flat. Once the luminal folds were exposed, the tops of the folds were removed and used for tissue extraction. Using the tops of these folds ensured that the most abundant cell type was the tubular gland cell.

Approximately 5 mg of each tissue type was used for genomic DNA isolation using a Qiagen Genomic DNA isolation kit. DNA was quantified and used in a PCR  
30 reaction with primers HC-1 and HC-4 that amplify a section of the human IgG heavy chain. The vector used for these injections served as a positive control in the PCR reactions. One LSU bird (Bird 2211) and one Bull Run bird (bird 2895) did not receive an injection and were used as negative controls.

In order to determine if gene incorporation (transposition) was occurring, instead of maintenance of the vector, PCR was conducted using a primer anchored in the gene of interest and the transposase. The result was a greatly reduced PCR reaction when compared to the PCR result from the heavy chain indicating the vector  
5 was being destroyed while the target gene was being maintained in the recipient chromosome.

PCR was conducted on the the liver of quail injected in the left ventricle with a transposon-based vector encoding for the monoclonal MCS(CMVp/pp/HC/ProLys/LC/CPA) SEQ ID NO:48. Primers designed to the  
10 heavy chain of the monoclonal resulted in the correct PCR fragment in all of the injected birds and the positive vector control. All control birds, PCR controls, and kit controls resulted in a negative PCR reaction. This PCR reaction proved DNA uptake by the liver in birds that received a cardiac injection. One bird was slightly weaker on an oviduct sample. These results clearly demonstrate DNA presence in high  
15 quantities two days post-cardiac injection.

In order to determine if transposition occurred, the same quantity of DNA was used in the PCR reaction containing primers HC 1 and HC 4 as was used in the PCR reaction containing primers mATS3'F and mATS5'R (these primers amplify a segment of DNA within the transposase). If no transposition occurred, then the bands  
20 in each reaction would be very similar in intensity. If transposition occurred, then the bands would not be similar. In order to make sure there was not a problem with the buffer chosen, 3 buffers from an optimization kit were used. Due to the band intensity from the initial PCR, the number of cycles was decreased from 45 to 30 in order to detect any small differences that might occur.

As seen in the transposition PCR, the band corresponding to the transposase was present, but at a concentration much less than the heavy chain fragment that was amplified. The results also demonstrate that the transposase is degraded which the gene encoding for the heavy chain is stably incorporated. This indicates that transposition has occurred and that the majority of the amplicon was due to copies  
25 integrated into the quail genome. Using such a delivery system combined with a transposon-based vector allows rapid expression of a gene for protein production in the liver or oviduct, or allows production of transgenic hens and roosters equivalent to G2 offspring if a traditional route of transfecting one animal and crossing to increase gene copy number is used.  
30

## EXAMPLE 7

*Intracardiac Injection of a Transposon-based Vector for Gene Therapy and Production of Proinsulin in Transgenic Chickens*

A total of 6 mature white leghorn hens were anesthetized and injected into the  
5 left ventricle of the heart with 1 ml total volume (consisting 50 µg of DNA, 150 µl of  
Superfect supplemented with HBSS to 1 ml). The methods are similar to those  
described in the preceding example for quail.

Briefly, a needle approximately 1 inch (22 gauge) in length was used for  
injections of chickens. The needles were connected to a 1ml tuberculin syringe  
10 containing the transfection mixture. Birds were held at the base of the wings on a  
table. Feathers in the area of the feather track about halfway down the breast were  
grasped and a few down feathers removed over the injection site. The area sprayed  
with ethanol.

The injector placed his left hand over the bird with the tip of the forefinger  
15 placed on the anterior tip of the keel. The thumb was used to palpate the triangle-  
shaped, posterior end of the caudolateral process of the sternum. The caudolateral  
process was followed forward to where it joined the body of the sternum. This  
marked the U shaped bony border of the lateral notch. The U of the lateral notch is  
formed by the thoracic process and the caudolateral process. While maintaining the  
20 thumb in the lateral notch, an imaginary line was drawn straight down from the  
forefinger. Another imaginary line was drawn at the angle of the caudolateral process  
from the tip of the thumb forward. The site where the needle was placed was the  
intersection of these lines. This is approximately 2cm towards the bird's head from  
the tip of the thumb.

25 The needle and syringe were held parallel to the table. The needle was  
inserted into the superficial pectoralis muscle. Without completely withdrawing the  
needle, it was repositioned slightly to one side or the other until an intercostal space  
was found.

The needle was placed into the left breast muscle at about a 45° angle. When  
30 the needle was about halfway in, the needle hit the sternum. Next, the needle was  
partially removed and repositioned at a steeper angle until the sternum was no longer  
encountered. At this angle the needle dropped into the left ventricle of the heart. A  
flash of blood appeared in the syringe and pulsed in the hub at the rate of the  
heartbeat. The plunger on the syringe was slowly depressed. If there was an air



bubble above the solution inside the syringe, the plunger was stopped before the air was pushed out into the blood. The needle was removed and disposed in a biohazard sharps container. The bird was returned to its cage and monitored for any signs of distress. (See A Color Atlas of Avian Anatomy. John McLelland. W.B. Saunders  
 5 Company, 1991 for view of the anatomy of this area).

The vector SEQ ID NO: 49 encoded for chicken ovalbumin::ent Tag::proinsulin fusion protein (Vector: pTnMCS (ChOVep/OVg'/ent/pro-ins/syn poly A) (Clone MCS6). Twenty-four hours post-injection, two birds were sacrificed and liver, ovary and oviduct tissue was removed from each bird. Genomic DNA was  
 10 extracted from each tissue as described previously. PCR was conducted and a sample of that reaction was electrophoresed on a 2% gel. The remaining 4 chickens are laying eggs that are currently being evaluated for the presence of the fusion protein.

PCR was conducted on DNA isolated from the liver, oviduct and ovary of two chickens (2004, 2005) injected in the left ventricle with a transposon-based vector  
 15 encoding for ovalbumin::ent Tag::proinsulin fusion protein SEQ ID NO:49. The result was a positive PCR reaction from each DNA sample, regardless of the tissue type and the vector control. All kit and PCR controls were negative indicating no contamination had occurred. The results show band amplified by PCR that indicate that the gene encoding for proinsulin is present in the liver, ovary and oviduct of each  
 20 of the two chickens examined.

SEQ ID NO: 49 pTnMCS(Chicken OVep+OVg'+ENT+proins+syn polyA) was constructed as follows:

Bp 1 – 3670 from-vector pTnMCS, bp 1 - 3670

25 Bp 3676 – 4350 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1 bp 1-675

Bp 4357 – 5692 Chicken Ovalbumin promoter taken from GenBank accession # J00895-M24999 bp 1-1336

Bp 5699 – 6917 Chicken Ovalbumin gene from GenBank Accession # V00383.1 bp  
 30 2- 1220. (This sequence includes the 5'UTR, containing putative cap site, bp 5699-5762.)

Bp 6924 - 7073 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site

Bp 7074 - 7334 Human proinsulin GenBank Accession # NM000207 bp 117-377

- Bp 7335 - 7379 Spacer DNA, derived as an artifact from the cloning vectors  
pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)  
Bp 7380 - 7731 Synthetic polyA from the cloning vector gWIZ (Gene Therapy  
Systems) bp 1920 - 2271  
5 Bp 7733 - 11332 from vector pTnMCS, bp 3716 - 7315

All patents, publications and abstracts cited above are incorporated herein by  
reference in their entirety, including U.S. provisional patent applications serial  
numbers 60/532,504, 60/565,371 and 60/592,098, and PCT patent applications  
10 PCT/US03/41261, PCT/US03/41269, and PCT/US03/41335. It should be understood  
that the foregoing relates only to preferred embodiments of the present invention and  
that numerous modifications or alterations may be made therein without departing  
from the spirit and the scope of the present invention as defined in the following  
claims.

15

TABLE 1

<b>Reproductive tissue</b>	<b>Promoter</b>	<b>Ref.</b>	<b>Function/comments</b>
testes, spermatogenesis	SPATA4	1	constitutive 30 d after birth in rat
placenta, glycoprotein	ERVWE1	2	URE, Upstream Regulatory Element is tissue spec. enhancer
breast epithelium and breast cancer	mammaglobin	6	specific to breast epithelium and cancer
prostate	EPSA	17	enhanced prostate-specific antigen promoter
testes	ATC	25	AlphaT-catenin specific for testes, skeletal, brain cardiomyocytes
prostate	PB	67	probasin promoter
<b>Vision</b>			
rod/cone	mCAR	3	cone photoreceptors and pinealocytes
retina	ATH5	15	functions in retinal ganglia and precursors
eye, brain	rhodopsin	27	
kertocytes	keratocan	42	specific to the corneal stroma
retina	RPE65	59	
<b>Muscle</b>			
vascular smooth muscle	TFPI	13	Tissue Factor Pathway Inhibitor – low level expression in endothelial and smooth muscle cells of vascular system
cardiac specific	MLC2v	14, 26	ventricular myosin light chain
cardiac	CAR3	18	BMP response element that directs cardiac specific expression
skeletal	C5-12	22	high level, muscle spec expression to drive target gene
skeletal	AdmDys, AdmCTLA4lg	32	muscle creatine kinase promoter
smooth muscle	PDE5A	41	chromosome 4q26, phosphodiesterase use intronic splicing elements to restrict expression to smooth muscle vs skeletal
smooth muscle	AlphaTM	45	
skeletal	myostatin	48	fiber type-specific expression of myostatin
<b>Endocrine/nervous</b>			
glucocorticoid	GR 1B-1E	4, 12	glucocorticoid receptor promoter/ all cells
neuroblastoma	M2-2	8, 36	M2 muscarinic receptor amyloid beta-protein; 30 bp fragment needed for PC12 and glial cell expression
brain	Abeta	16	neuron-specific; high in hippocampus,
brain	enolase	21	intermediate in cortex, low in cerebellum clusters acetylcholine receptors at
synapses	rapsyn	29	neuromuscular junction express limited to neurons in central and peripheral nervous system and specific endocrine cells in adenohypophysis,
neuropeptide precursor	VGF	39	adrenal medulla, GI tract and pancreas use of methylation to control tissue specificity in neural cells.
mammalian nervous system central and peripheral noradrenergic neurons	BMP/RA Phox2a/Phox2b	46 47	regulation of neuron differentiation

brain	BAI1-AP4	55	spec to cerebral cortex and hippocampus
<b>Gastrointestinal</b>			
UDP glucoronsyltransferase	UGT1A7	11	gastric mucosa
	UGT1A8	11	small intestine and colon
	UGT1A10	11	small intestine and colon
colon cancer	PKCbetall	20	Protein kinase C betall (PKCbetall); express in colon cancer to selectively kill it.
<b>Cancer</b>			
tumor suppressor 4.1B	4.1B	5	2 isoforms, 1 spec to brain, 1 in kidney
nestin	nestin	63	second intron regulates tissue specificity
cancer spec promoter	hTRT/hSPA1	68	dual promoter system for cancer specificity
<b>Blood/lymph system</b>			
Thyroid	thyroglobulin	10	Thyroid spec. -- express to kill thyroid tumors
Thyroid	calcitonin	10	medullary thyroid tumors
Thyroid	GR 1A	12	regulation controlled by DREAM
thyroid	thyroglobulin	50	transcriptional repressor
arterial endothelial cells	ALK1	60	activin receptor-like kinase
<b>Nonspecific</b>			
RNA polymerase II		7	
gene silencing	Gnasx1, Nespas	31	
beta-globin	beta globin	53	
<b>Cardiac</b>			
	M2-1	8	M2 muscarinic receptor
<b>Lung</b>			
	hBD-2	19	IL-17 induced transcription in airway epithelium
pulmonary surfactant protein	SP-C	62	Alveolar type II cells
ciliated cell-specific prom	FOZJ1	70	use in ciliated epithelial cells for CF treatment
surfactant protein expression	SPA-D	73	Possible treatment in premature babies
Clara cell secretory protein	CCSP	75	
<b>Dental</b>			
teeth/bone	DSPP	28	extracellular matrix protein dentin sialophosphoprotein
<b>Adipose</b>			
adipogenesis	EPAS1	33	endothelial PAS domain -- role in adipocyte differentiation
<b>Epidermal</b>			
differentiated epidermis	involucrin	38	
desmosomal protein	CDSN	58	stratum granulosum and stratum corneum of epidermis
<b>Liver</b>			
liver spec albumin	Albumin	49	
serum alpha-fetoprotein	AFP	56	liver spec regulation

## References

1. Biol Pharm Bull. 2004 Nov;27(11):1867-70
2. J Virol. 2004 Nov;78(22):12157-68
- 5 3. Invest Ophthalmol Vis Sci. 2004 Nov;45(11):3877-84
4. Biochim Biophys Acta. 2004 Oct 21;1680(2):114-28
5. Biochim Biophys Acta. 2004 Oct 21;1680(2):71-82
6. Curr Cancer Drug Targets. 2004 Sep;4(6):531-42
7. Biotechnol Bioeng. 2004 Nov 20;88(4):417-25
- 10 8. J Neurochem. 2004 Oct;91(1):88-98
10. Curr Drug Targets Immune Endocr Metabol Disord. 2004 Sep;4(3):235-44
11. Toxicol Appl Pharmacol. 2004 Sep 15;199(3):354-63
12. J Immunol. 2004 Sep 15;173(6):3816-24
13. Thromb Haemost. 2004 Sep;92(3):495-502
- 15 14. Acad Radiol. 2004 Sep;11(9):1022-8
15. Development. 2004 Sep;131(18):4447-54
16. J Neurochem. 2004 Sep;90(6):1432-44
17. Mol Ther. 2004 Sep;10(3):545-52
18. Development. 2004 Oct;131(19):4709-23. Epub 2004 Aug 25
- 20 19. J Immunol. 2004 Sep 1;173(5):3482-91
20. J Biol Chem. 2004 Oct 29;279(44):45556-63. Epub 2004 Aug 20
21. J Biol Chem. 2004 Oct 22;279(43):44795-801. Epub 2004 Aug 20
22. Hum Gene Ther. 2004 Aug;15(8):783-92
25. Nucleic Acids Res. 2004 Aug 09;32(14):4155-65. Print 2004
- 25 26. Mol Imaging. 2004 Apr;3(2):69-75
27. J Gene Med. 2004 Aug;6(8):906-12
28. J Biol Chem. 2004 Oct 1;279(40):42182-91. Epub 2004 Jul 28
29. Mol Cell Biol. 2004 Aug;24(16):7188-96
31. Nat Genet. 2004 Aug;36(8):894-9. Epub 2004 Jul 25
- 30 32. Gene Ther. 2004 Oct;11(19):1453-61
33. J Biol Chem. 2004 Sep 24;279(39):40946-53. Epub 2004 Jul 15
36. Brain Res Mol Brain Res. 2004 Jul 26;126(2):173-80
38. J Invest Dermatol. 2004 Aug;123(2):313-8
39. Cell Mol Neurobiol. 2004 Aug;24(4):517-33
- 35 41. Int J Impot Res. 2004 Jun;16 Suppl 1:S8-S10
42. Invest Ophthalmol Vis Sci. 2004 Jul;45(7):2194-200
45. J Biol Chem. 2004 Aug 27;279(35):36660-9. Epub 2004 Jun 11
46. Brain Res Mol Brain Res. 2004 Jun 18;125(1-2):47-59
47. Brain Res Mol Brain Res. 2004 Jun 18;125(1-2):29-39
- 40 48. Am J Physiol Cell Physiol. 2004 Oct;287(4):C1031-40. Epub 2004 Jun 09
49. Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi. 2003 Nov;19(6):601-3
50. J Biol Chem. 2004 Aug 6;279(32):33114-22. Epub 2004 Jun 04
53. Brief Funct Genomic Proteomic. 2004 Feb;2(4):344-54
55. FEBS Lett. 2004 May 21;566(1-3):87-94
- 45 56. Biochem Biophys Res Commun. 2004 Jun 4;318(3):773-85
58. J Invest Dermatol. 2004 Mar;122(3):730-8
59. Mol Vis. 2004 Mar 26;10:208-14
60. Circ Res. 2004 Apr 30;94(8):e72-7. Epub 2004 Apr 01
62. Am J Physiol Lung Cell Mol Physiol. 2004 Dec 3; [Epub ahead of print]
- 50 63. Lab Invest. 2004 Dec;84(12):1581-92
67. Prostate. 2004 Jun 1;59(4):370-82
68. Cancer Res. 2004 Jan 1;64(1):363-9
70. Mol Ther. 2003 Oct;8(4):637-45
73. Front Biosci. 2003 May 01;8:d751-64

75. Am J Respir Cell Mol Biol. 2002 Aug;27(2):186-93

Table 2

Gene-protein target*	Cellular function	Type of cancer tested
<i>B-raf</i>	Serine/threonine kinase	Malignant melanoma
Nox1	Superoxide-generating oxidase	Transformed MRK cells <sup>a</sup>
FAS/Her2	Fatty acid synthase	Breast-MDA-MB-231
Cyclin E	Cell-cycle control	Hepatocarcinoma
Hec1	Chromosomal segregation	
Gp210	Nuclear pore assembly	Adenocarcinoma (Hela cells)
c-Kit	Signal transduction	Gastrointestinal
MDR	Multi-drug resistance	Adenocarcinoma (Hela cells)
<i>bcl-2</i>	Antiapoptotic	Esophageal adenocarcinoma
<i>livin</i>	Antiapoptotic	Adenocarcinoma
<i>survivin</i>	Antiapoptotic	Adenocarcinoma (Hela cells)
Philadelphia chromosome		Chronic myeloid leukemia
Ribonucleotide reductase	Gemcitabine resistance	Hepatic metastasis
Rho C	Cell motility	Metastasis

<sup>a</sup>Normal rat kidney cells.

\*Genes are written in italics and lower case letters while proteins begin with a capital letter and are written in roman letters.

Table 3

<b>Genetic Disorder</b>	<b>Gene</b>	<b>Chromosome</b>
Hemophilia		
Depression	CRHR1	
SCIDS	SCID-X1	X
Breast Cancer	ESR1	
	BRCA1	
	BRCA2	
	p53	
Lupus (SLAM/C D2)	Sles1	
	Sles2	
	Sles3	
	Sles4	
Colon Cancer	15-PGDH	
	MSH2	2
	MSH6	2
	MLH1	3
Crohn's Disease	CD 19	16
	sialophorin	
	CD11 integrin	
	IL-4	
Cystic Fibrosis	CFTR	
Type I Diabetes	IDDM1	6
	IDDM2	11
	GCK (glucokinase)	7
Glucose/Galactose Malabsorption (CGM)	SGLT1	22
Pancreatic Cancer	DPC4 (Smad4)	18
	p53	
	Rb	
Wilson's disease	ATP7B	13
Zellweger syndrome	PXR 1	12
Sickle Cell Anemia	HBB	11p15.4
Burkitt Lymphoma	Myc	8
Gaucher disease	glucocerebroside	
Hemophilia A	HEMA (Factor VIII)	X
Chronic Myeloid leukemia	ABL (9) BCR (22)	9/22 exchange
Niemann-Pick Type A, B or C	NP-C	18
Hemoglobinuria (PNH)	PIG-A	X
Porphyria		
Thalassemia alpha	HBA1	16
	HBA2	16
Thalassemia beta	HBB	11
Small cell lung carcinoma		3
Melanoma	CDKN2	9
Multiple endocrine neoplasia	MEN1	11
Neurofibromatosis	NF-2	22
Li-Fraumeni syndrome	p53	17
	Rb	13
Polycystic Kidney Disease	PKD1	16
Prostate cancer	HPC1	1
Harvey Ras oncogene	Ras	11
Tuberous sclerosis	TSC1	9
	TSC2	16



Von-Hippel Lindau	VHL	3
Deafness	Cx26	13q11-12
Pendred syndrome	PDS	7
Best disease	VMD2	11
Glaucoma	GLC1A	1
gyrate atrophy	OAT	10
Rett syndrome	MeCP2	Xq28
Congenital adrenal hyperplasia	CYP21P	6
Adrenaleukodystrophy	ALD	
AI polyglandular syndromel	AIRE	21
Cockayne syndrome Type I	CSA	5
Cockayne syndrome Type II	CSB	
Diastrophic dysplasia	DTD	5
Ataxia telangiectasia	ATM	11
Atherosclerosis	Apo E	19
Long QT syndrome	LQT1	11
Williams Syndrome	LIM kinase, elastin	7
Asthma		5,6, 11, 14, 12
DiGeorge syndrome		22
Hyper-IgM	TNFSF5	Xq26
Severe Combined Immunodeficiency Disease (SCID)	IL2RG	X
	JAK3	19
	ADA	20
Alport Syndrome	COL4A5	X
5-alpha reductase		5
Achondroplasia	FGFR3	4
Familial ALS	SOD1	21
Charcot-Marie-Tooth disease Type 1A	PMP22	17
Charcot-Marie-Tooth disease Type 1B		X
Dejerine-Sottas Syndrome		
Duchenne muscular dystrophy	dystrophin	X
Ellis-van Creveld syndrome	EVC	4
Fibrodysplasia Ossificans Progressiva	NOG	
	BMP	
Marfan Syndrome	FBN1	15
Myotonic dystrophy	myotonic dystrophy	19
Fragile X	FMR1	X
PWS	SNRPN	15
Waardenburg syndrome	Pax3	2
Werner disease	SGS1	8
Alzheimer disease	PS1	14
	PS2	1
Angelman syndrome	deletion 15q11q13	15
	UBE3A	
Essential tremor	ETM1	3
	ETM2	2
Familial Mediterranean fever	FMF	16
Friedreich's ataxia	YFH1	
Huntington's disease	HD	4
Maple Syrup Urine Disease	BCKDH complex	
Parkinson's	Alpha-synuclein	4
Refsum disease	PAHX	10
Spinal Muscular Atrophy	SMN1	5
	SMN2	
spinocerebellar ataxia	SCA1	6

Tangier Disease  
Tay-Sach's  
Obesity  
AAT  
Hemochromatosis

ABC1  
HEXA  
leptin  
SERPINA 1  
HFE (mutation C282Y)  
HFW (mutation H63D)

9q31  
15  
7  
14

## CLAIMS

We claim:

- 5           1.       A composition comprising a transposon-based vector comprising :
- a)       a gene operably linked to a first promoter, the gene encoding  
                  for a bacterial transposase; and,
- b)       one or more genes of interest operably-linked to one or more  
                  additional promoters,
- 10                wherein the one or more genes of interest and their operably-linked  
                  promoters are flanked by transposase insertion sequences recognized by the  
                  bacterial transposase, wherein the first promoter and the one or more  
                  additional promoters are cell-specific promoters or constitutive promoters.
- 15           2.       The transposon-based vector of claim 1, further comprising an isolated  
                  polyA nucleotide sequence located 3' to the one or more genes of interest.
3.       The isolated polyA nucleotide sequence of claim 2, wherein the  
                  isolated polyA nucleotide sequence is optimized for production of a protein,  
20                peptide or nucleic acid encoded by the one or more genes of interest.
4.       The transposon-based vector of claim 1, wherein the one or more genes  
                  of interest code for a protein, a peptide or a nucleic acid.
- 25           5.       The transposon-based vector of claim 1, wherein the one or more gene  
                  of interest encodes for a nucleic acid which inhibits transcription.
6.       A composition comprising an isolated polynucleotide sequence  
                  comprising:
- 30                a) one or more genes of interest operably-linked to one or more  
                  promoters;
- b) a poly A nucleotide sequence located 3' to the one or more genes of  
                  interest; and,

c) transposase insertion sequences recognized by a bacterial transposase,

wherein the one or more genes of interest and their operably-linked promoters are flanked by the transposase insertion sequences and the one or more additional promoters are cell-specific promoters or constitutive promoters.

7. The isolated polynucleotide sequence of claim 6, wherein the one or more genes of interest code for a protein, a peptide or a nucleic acid.

8. An animal or a human comprising the isolated polynucleotide sequence of claim 6.

9. The animal of claim 8, wherein the animal is a bird or a mammal.

10. An egg produced by the bird of claim 9.

11. Milk produced by the mammal of claim 9.

12. A cell comprising the isolated polynucleotide sequence of claim 6.

13. A method of providing gene therapy to an animal or a human comprising administering to the animal or the human the transposon-based vector of Claim 1.

14. The method of claim 13, wherein the one or more additional promoter is a cell specific promoter.

15. The method of claim 13, wherein the gene of interest codes for production of a protein, peptide or nucleic acid.

16. The method of claim 13, further comprising a polyA sequence located 3' to the one or more genes of interest.

17. The method of claim 13, wherein the gene therapy comprises production of a protein, peptide or nucleic acid encoded by the one or more genes of interest in the animal or the human.
- 5 18. The method of claim 13, wherein the administration is effective to treat a disease or a condition.
19. The method of claim 13, wherein the administration of the transposon-based vector results in a transfection rate of at least 40%.
- 10 20. The method of claim 13, wherein the administration occurs through the vascular system.
21. An animal produced by the method of claim 13.
- 15 22. Use of the composition of any one of claims 1-7, in the preparation of a medicament useful for providing gene therapy to an animal or human following administration of an effective amount of the composition to the animal or the human.
- 20 23. The use of claim 22, wherein the gene therapy treats a disease or a condition in the animal or the human.

FIGURE 1

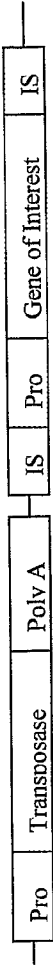


FIGURE 2

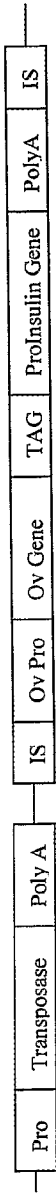


FIGURE 3

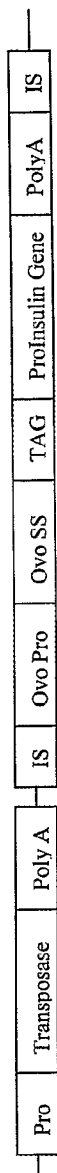




FIGURE 4

IS	Tet <sub>i</sub> Pro	Ovgen	Pro	Ovotrans	Pro	Ovomucin	IS
----	----------------------	-------	-----	----------	-----	----------	----

Figure 5.

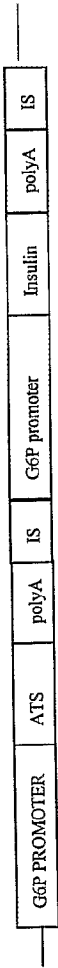


Figure 6

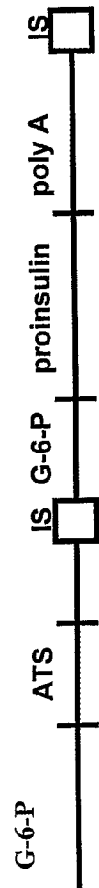


Figure 7



Figure 8

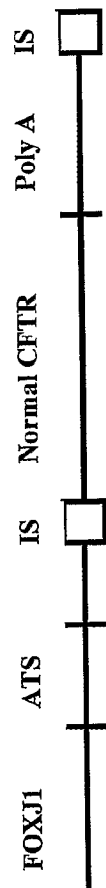




Figure 9

## APPENDIX

SEQ ID NO:1  
 5 atg ctg ggc atc tgg acc ctc cta cct ctg gtt ctt acg tct gtt gct aga tta

SEQ ID NO:2 (cecropin pro)  
 GCG CCA GAG CCG AAA

10 SEQ ID NO:3(cecropin pro extended)  
 GCG CCA GAG CCG AAA TGG AAA GTC TTC AAG

SEQ ID NO:4 (cecropin prepro)  
 15 AAT TTC TCA AGG ATA TTT TTC TTC GTG TTC GCT TTG GTT CTG GCT TTG TCA ACA  
 GTT TCG GCT GCG CCA GAG CCG AAA

SEQ ID NO:5 (cecropin prepro extended)  
 20 AAT TTC TCA AGG ATA TTT TTC TTC GTG TTC GCT TTG GTT CTG GCT TTG TCA ACA  
 GTT TCG GCT GCG CCA GAG CCG AAA TGG AAA GTC TTC AAG

SEQ ID NO:6 (pTnMCS)

1 ctgacgcgcc ctgtagcgcc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga  
 25 61 ccgctacact tgccagcgcc ctacgcgcc ctcccttcgc tttcttccct tcccttctcg  
 121 ccacgttcgc cggcgcacga ttgctatttg gccattgcat acgttgatc catatcataa  
 181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac  
 241 tagttattaa tagtaataa ttacggggtc attagtcat agcccatata tggagttccg  
 301 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgccatt  
 361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca  
 421 atgggttgag tatttacggg aaactgocca cttggcagta catcaagtgt atcatatgcc  
 481 aagtagcccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgccagta  
 541 catgacctta tgggactttc ctacttgcca gtacatctac gtattagtoa tgcctattac  
 601 catgggtgatg cggttttggc agtacatcaa tggcgtgga tagcgggttg actcacgggg  
 35 661 atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg  
 721 ggactttcca aaatgtcgt acaactccgc cccattgacg caaatggcg gtaggcgtgt  
 781 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg  
 841 ccattccagc tgttttgacc tccatagaag acaccgggac cgtatccagcc tccgcgccg  
 901 ggaacgggtg attggaacgc ggattcccg tgccaagagt gacgtaagta ccgcctatag  
 40 961 actctatagg cacaccctt tggctcttat gcatgctata ctgtttttgg cttggggcct  
 1021 atacaccccc gcttccttat gctatagggtg atgggtatagc ttagcctata ggtgtgggtt  
 1081 attgaccatt attgaccact ccctatttgg tgacgataact ttccattact aatccataac  
 1141 atggctcttt gccacaacta tctctatttg ctatatgcca atactctgtc cttcagagac  
 1201 tgacacggac totgtatttt tacaggatgg ggtccatttt attatttaca aattcacata  
 45 1261 tacaacaacg ccgtcccccg tgcccgcagt ttttattaaa catagcgtgg gatctccacg  
 1321 cgaatctcgg gtacgtgttc cggacatggg ctcttctccg gtacggcgcg agcttccaca  
 1381 tccgagccct ggtcccatgc ctccagcggc tcatggtcgc tccgagctc cttgtctcta  
 1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag  
 1501 gccgtgcccg tagggtagt gtctgaaaat gagcgtggag attgggctcg caccgctgac  
 50 1561 gcagatggaa gacttaaggg agcggcagaa gaagatgcag gcagctgagt tgttgtattc  
 1621 tgataagagt cagaggtaac tcccgttgcg gtgctgttaa cgttgagggg cagtgtatgc  
 1681 tgagcagtac tegtgtctgc cgcgcgcgcc accagacata atagctgaca gactaacaga  
 1741 ctgttccctt ccatgggtct tttctgcagt caccgtcgga ccatgtgcga actcgatatt  
 1801 ttacacgact ctctttacca attctgcccc gaattacact taaaacgact caacagctta  
 55 1861 acgtttgctt gccacgcat acttgactgt aaaactctca ctcttaccga acttggccgt  
 1921 aacctgcca ccaaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt  
 1981 aatcgtcacc tccacaaaga gcgactcgct gtataccgtt ggcatgctag ctttatctgt  
 2041 tccggcaata cgtatcccat tgtacttgt gactggtctg atattcgtga gcaaaaacga  
 2101 cttatggtat tgcgagcttc agtcgcacta caccgtcgtt ctgttactct ttatgagaaa  
 60 2161 gcgttcccg cttcagagca atgttcaaag aaagctcatg accaatttct agccgacctt  
 2221 gcgagcattc taccgagtaa caccacaccg ctcatgttca gtgatgctgg ctttaaagtg  
 2281 ccatggtata aatccgttga gaagctgggt tggtagctgt taagtcgagt aagagggaaa  
 2341 gtacaatatg cagacctagg agcggaacac tggaaccta tcagcaactt acatgatag  
 2401 tcactctagc actcaaaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca

	2461	tgccaaattc	tattgtataa	atctcgcctc	aaaggccgaa	aaaatcagcg	ctcgacacgg
	2521	actcattgtc	accaccgcgc	acctaataatc	tactcagcgt	cggcaaaagg	gccatggggt
	2581	ctagcaacta	acttacctgt	tgaaattcga	acacccaaac	aacttggtta	tatctattcg
5	2641	aagcgaatgc	agattgaaga	aaccttcoga	gacttgaaaa	gtcctgccta	cggactaggg
	2701	ctacgccata	gccgaacgag	cagctcagag	cgttttgata	tcatgctgct	aalcgcctcg
	2761	atgcttcaac	taacatgttg	gottgcgggc	gttcattgctc	agaaacaagg	ttgggacaag
	2821	cacttccagg	ctaaccacgt	cagaaatcga	aacgtactct	caacagttcg	cttaggcattg
	2881	gaagttttgc	ggcattctgg	ctacacaata	acaaggggaag	acttactcgt	ggctgcaacc
	2941	ctactagctc	aaaattttatt	cacacatggg	tacgctttgg	ggaaattatg	aggggatcgc
10	3001	tctagagcga	tccgggatct	cgggaaaagc	gttggtgacc	aaaggtgcct	tttatcatca
	3061	ctttaaaaat	aaaaaacaat	tactcagtcg	ctgttataag	cagcaattaa	ttatgattga
	3121	tgccctacatc	acaacaaaaa	ctgatttaac	aaatggttgg	tctgccttag	aaagtatat
	3181	tgaacattat	cttgaltata	ttattgataa	taataaaaaa	cttatcccta	tccaagaaat
	3241	gatgcctatc	attgggttga	atgaacttga	aaaaaattag	ccttgaatac	attactggta
15	3301	aggtaaacgc	cattgtcagc	aaattgatcc	aagagaacca	acttaagct	ttcctgacgg
	3361	aatgttaatt	ctcgttgacc	ctgagcactg	atgaatcccc	taatgatttt	ggtaaaaaatc
	3421	attaagttaa	gggtgatata	catcttgctc	tatgatcccg	gtaatgtgag	ttagctcaat
	3481	cattaggcac	cccaggcttt	acactttatg	cttcgggctc	gtatgtttgtg	tggaattgtg
	3541	agcggataac	aatttcacac	aggaaaacgc	tatgaccatg	attacgccaa	gcgcgcaatt
20	3601	aaccctcaat	aaagggaaca	aaagctggag	ctccaccgcg	gtggcggcgg	ctctagaact
	3661	agtggtatcc	ccgggctgca	ggaattcgat	atcaagctta	tcgataccgc	tgacctcgag
	3721	ggggggcccg	gtacccaatt	cgccctatag	tgagtctgat	taocgcgcgt	cactggccgt
	3781	cgttttacaa	cgctgtgact	gggaaaaacc	tgccgttacc	caacttaata	gccttgacgc
	3841	acatccccc	ttcgccagct	ggcgttaata	cgaagaggcc	cgcaccgata	gcctttacca
25	3901	acagttgcgc	agcctgaatg	gcgaatggaa	atttgaagcg	ttaatatttt	gttaaaatc
	3961	gcgttaaat	tttggttaaat	cagctcattt	tttaaccaat	aggccgaaat	cggcaaaatc
	4021	ccttataaat	caaaagaata	gacgagata	gggttgagtg	ttgttccagt	ttggaacaag
	4081	agtcactact	taaaagacgt	ggactccaac	gtcaaaaggc	gaaaaaccgt	ctatcagggc
	4141	gatggccac	tactccggga	tcatatgaca	agatgtgtat	ccaccttaac	ttaatgattt
30	4201	ttaccaaaat	cattagggga	ttcatcagtg	ctcagggtca	acgagaatta	acattccgct
	4261	aggaaagcct	atgatgatga	tgtgcttaaa	aacttactca	atggctgggt	atgcatatcg
	4321	caatacatgc	gaaaaaccata	aaagagcttg	ccgataaaaa	aggccaattt	attgctattt
	4381	accgcccgtt	tttattgagc	ttgaaagata	aataaaatag	atagggttta	tttgagcta
	4441	aatcttcttt	atcgtaaaaa	atgccctctt	gggttatcaa	gagggtcatt	atatttcgcg
35	4501	gaataacatc	atttgggtgac	gaaataacta	agcacttgct	tcctgtttac	tccctgagc
	4561	ttgaggggtt	aacatgaagg	tcatcgatag	caggataata	atacagtaaa	acgctaaacc
	4621	aataatccaa	atccagccat	cccaaattgg	tagtgaatga	ttataaataa	cagcaaacag
	4681	taattggcca	ataacaccgg	ttgcattggt	aaggctcacc	aataatccct	gtaaagcacc
	4741	ttgctgatga	ctctttgttt	ggatagacat	cactccctgt	aatgcaggta	aagcgatccc
40	4801	accaccagcc	aataaaatta	aaacaggga	aactaaccaa	ccttcagata	taaacgctaa
	4861	aaaggcaaat	gcactactat	ctgcaataaa	tccgagcagt	actgcccgtt	tttcgcccac
	4921	ttagtggtga	ttcttccctg	cacaaaggct	tgaataactg	agtgtaaaaa	accaagaccc
	4981	gtaatgaaaa	gccaaccatc	atgctattca	tcatacacgt	ttctgttaata	gcaccacacc
	5041	gtgctggatt	ggctatcaat	gcgctgaat	aataatcaac	aaatggcatc	gttaataaag
45	5101	tgatgtatc	cgatcagctt	ttgttccctt	tagtgagggt	taattgcggg	cttgccgttaa
	5161	tcattggtat	agctgtttcc	tgtgtgaaat	tggttatccg	tcacaattcc	acacacata
	5221	cgagccggaa	gcataaagtg	taaaagcctg	gggtgcctaat	gagtgcgcta	actcacatta
	5281	attgcgttgc	gctcactgoc	cgctttccag	tcgggaaacc	tgtcgtgcca	gctgcattaa
	5341	tgaatcgccc	aacgcgcggg	gagaggcggt	ttgcgtattg	ggcgtctctc	cgcttccctg
50	5401	ctcactgact	cgctgcgcct	ggctgcttcg	ctgcggcgag	cggtatcagc	tcactcaaa
	5461	gcggtaatac	ggttatccac	agaatcaggg	gataacgcag	gaaagaacat	gtgagcaaaa
	5521	ggccagcaaa	aggccaggaa	ccgtaaaaag	gccgcgttgc	tggcgttttt	ccataggctc
	5581	cgcccccttg	acgagcatca	caaaaatcga	cgctcaagtc	agaggtggcg	aaacccgaca
	5641	ggactataaa	gataccaggg	gtttccccc	ggaagctccc	tcgtgcgctc	tcctgttccg
55	5701	accctgccc	ttaccggata	cctgtccgcc	ttttccctt	cgggaagcgt	ggcgttttct
	5761	catagctcac	gctgtaggta	tctcagttcg	gtgtaggctg	ttcgtcccaa	gctgggctgt
	5821	gtgcacgaac	ccccggttca	gcccgaacgc	tgccgcttat	ccggtaacta	togctttgag
	5881	tccaaccggg	taagacacga	cttatcgcca	ctggcagcag	ccactggtaa	caggattagc
	5941	agagcgaggt	atgtaggcgg	tgctacagag	ttcttgaagt	gggtggccta	ctacggctac
60	6001	actagaagga	cagtattttg	tatctgcgct	ctgctgaagc	cagttacctt	cggaaaaaga
	6061	gttggttagct	cttgatccgg	caaacaaaac	accgctggta	gcggtgggtt	ttttgtttgc
	6121	aagcagcaga	ttacgcgcag	aaaaaaagga	tctcaagaag	atcctttgat	cttttctacg
	6181	gggtctgacg	ctcagtgga	cgaaaactca	cgtaaaggga	ttttggctcat	gagattatca
	6241	aaaaggatct	tcacctagat	ccttttaaat	taaaaaatgaa	gttttaaatc	aatctaaagt
65	6301	atataatgag	aaacttggtc	tgacagttac	caatgcttaa	tcagtggagg	acctatctca
	6361	gcgatctgtc	tatttctgtc	atccatagtt	gcctgactcc	ccgtcgtgta	gataactacg
	6421	atacgggagg	gcttaccatc	tggccccagt	gctgcaatga	taccgcgaga	cccacgctca
	6481	ccggctccag	atttatcagc	aataaaaccag	ccagccggaa	gggcgcgagc	cagaagtggg



6541 cctgcaactt tatccgcctc catccagtct attaattgtt gccgggaagc tagagtaagt  
 6601 agttccgcag ttaatagttt gcgcaacggt gttgccattg ctacaggcat cgtggtgtca  
 6661 cgctcgtcgt ttggtagtgc ttcatcagc tccggttccc aacgatcaag gcgagttaca  
 6721 tgatccccc tgttgtgcaa aaaagcggtt agctccttcg gtccctccgat cgttgtcaga  
 5 6781 agtaagttgg ccgcagtggt atcactcatg gttatggcag cactgcataa ttctcttact  
 6841 gtcatgccat ccgtaagatg cttttctgtg actggtgagt actcaaccaa gtcattctga  
 6901 gaatagtgta tgcggcgacc gagttgctct tgcccggcgt caatacggga taataccgcy  
 6961 ccacatagca gaactttaa agtgctcatc attggaaaac gtcttcggg gcgaaaactc  
 7021 tcaaggatct taccgctggt gagatccagt tcatgtaac ccactcgtgc acccaactga  
 10 7081 tcttcagcat cttttacttt caccagcgtt tctgggtgag caaaaacagg aaggcaaat  
 7141 gccgcaaaaa aggggaataag ggcgacacgg aaatgttgaa tactcactact cttccttttt  
 7201 caatattatt gaagcattta tcagggttat tgtctcatga gcggatacat atttgaatgt  
 7261 atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt gccac

15

SEQ ID NO:7 (pTnMod)

CTGACGCGCC CTGTAGCGGC GCATTAAGCG CGGCGGGTGT GGTGGTTACG 50  
 CGCAGCGTGA CCGCTACACT TGCCAGCGCC CTAGCGCCCG CTCCTTTCGC 100  
 20 TTTCTTCCCT TCCTTTCTCG CCACGTTTCG CGGCATCAGA TTGGCTATTG 150  
 GCCATTGCAT ACGTTGTATC CATATCATAA TATGTACATT TATATTGGCT 200  
 CATGTCCAAC ATTACCGCCA TGTTGACATT GATTATTGAC TAGTTATTAA 250  
 TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA TGGAGTTCG 300  
 CGTTACATAA CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC 350  
 25 CCCGCCCATT GACGTCAATA ATGACGTATG TTCCCATAGT AACGCCAATA 400  
 GGGACTTTCC ATTGACGTCA ATGGGTGGAG TATTTACGGT AAAC TGCCCA 450  
 CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCC CCTATTGACG 500  
 TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCCAGTA CATGACCTTA 550  
 TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC 600  
 30 CATGCTGATG CGGTTTGGC AGTACATCAA TGGGCGTGGA TAGCGGTTTG 650  
 ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTTG 700  
 TTTTGGCACC AAAATCAACG GGACTTTCCA AAATGTCTGA ACAACTCCGC 750  
 CCCATTGACG CAAATGGGCG GTAGCGTGT ACGGTGGGAG GTCTATATAA 800  
 GCAGAGCTCG TTTAGTGAAC CGTCAGATCG CCTGGAGACG CCATCCACGC 850  
 35 TGTTTTGACC TCCATAGAAG ACACCGGGAC CGATCCAGCC TCCGCGGCCG 900  
 GGAACGGTGC ATTGGAACGC GGATTCCCCG TGCCAAGAGT GACGTAAGTA 950  
 CCGCTATAG ACTCTATAGG CACACCCCTT TGGCTCTTAT GCATGCTATA 1000  
 CTGTTTTTGG CTTGGGGCCT ATACACCCCT GCTTCCTTAT GCTATAGGTG 1050  
 ATGGTATAGC TTAGCCTATA GGTGTGGGTT ATTGACCATT ATTGACCACT 1100  
 40 CCCCATTGG TGACGATACT TTCCATTACT AATCCATAAC ATGGCTCTTT 1150  
 GCCACAACCTA TCTCTATTGG CTATATGCCA ATACTCTGTC CTTGAGAGAC 1200  
 TGACACGGAC TCTGTATTTT TACAGGATGG GGTCCCATT ATTATTTACA 1250  
 AATTACATA TACAACAACG CCGTCCCCCG TGCCCGCAGT TTTTATTAAA 1300  
 CATAGCGTGG GATCTCCACG CGAATCTCGG GTACGTGTTC CGGACATGGG 1350  
 45 CTCTTCTCCG GTAGCGGCGG AGCTTCCACA TCCGAGCCCT GGTCCCATGC 1400  
 CTCCAGCGGC TCATGGTTCG TCGGCAGCTC CTTGCTCCTA ACAGTGGAGG 1450  
 CCAGACTTAG GCACAGCACA ATGCCACCA CCACCAGTG GTCCGACACAAG 1500  
 GCCGTGGCGG TAGGGTATGT GTCTGAAAAT GAGCGTGGAG ATTGGGCTCG 1550  
 CACGGCTGAC GCAGATGGAA GACTTAAGGC AGCGGCAGAA GAAGATGCAG 1600  
 50 GCAGCTGAGT TGTGTATTC TGATAAGAGT CAGAGGTAAC TCCCGTTGCG 1650  
 GTGCTGTAA CGGTGGAGGG CAGTGTAGTC TGAGCAGTAC TCGTTGCTGC 1700  
 CGCGCGCGCC ACCAGACATA ATAGCTGACA GACTAACAGA CTGTTCCCTT 1750  
 CCATGGGTCT TTTCTGCAGT CACCGTCGGA CCATGTGTGA ACTTGATATT 1800  
 TTACATGATT CTCTTTACCA ATTCTGCCCC GAATTACACT TAAAACGACT 1850  
 55 CAACAGCTTA ACGTTGGCTT GCCACGCATT ACTTGACTGT AAAACTCTCA 1900  
 CTCTTACCGA ACTTGGCCGT AACCTGCCAA CCAAAGCGAG AACAAAACAT 1950  
 AACATCAAAC GAATCGACCG ATTGTTAGGT AATCGTCACC TCCACAAAGA 2000  
 GCGACTCGCT GTATACCGTT GGCATGCTAG CTTTATCTGT TCGGGAATAC 2050  
 GATGCCCAT GTACTTGTG ACTGGTCTGA TATTCGTGAG CAAAACGAC 2100  
 60 TTATGGTATT CCGAGCTTCA GTCGCACTAC ACGGTCGTT TGTACTCTT 2150  
 TATGAGAAAG CGTCCCCTT TTCAGAGCAA TGTTCAAAGA AAGCTCATGA 2200  
 CCAATTCTA GCCGACCTTG CGAGCATTCT ACCGAGTAAC ACCACACCGC 2250

	TCATTGTCAG	TGATGCTGGC	TTTAAAGTGC	CATGGTATAA	ATCCGTTGAG	2300
	AAGCTGGGTT	GGTACTGGTT	AAGTCGAGTA	AGAGGAAAAG	TACAATATGC	2350
	AGACCTAGGA	GCGGAAACT	GGAAACCTAT	CAGCAACTTA	CATGATATGT	2400
	CATCTAGTCA	CTCAAAGACT	TTAGGCTATA	AGAGGCTGAC	TAAAAGCAAT	2450
5	CCAATCTCAT	GCCAAATCT	ATTGTATAAA	TCTCGCTCTA	AAGGCCGAAA	2500
	AAATCAGCGC	TCGACACGGA	CTCATTGTCA	CCACCCGTCA	CCTAAAATCT	2550
	ACTCAGCGTC	GGCAAAGGAG	CCATGGGTTT	TAGCAACTAA	CTTACCTGTT	2600
	GAAATTCGAA	CACCCAAACA	ACTTGTTAAT	ATCTATTGCA	AGCGAATGCA	2650
	GATTGAAGAA	ACCTTCCGAG	ACTTGAAAAG	TCCTGCCTAC	GGACTAGGCC	2700
10	TACGCCATAG	CCGAACGAGC	AGCTCAGAGC	GTTTTGATAT	CATGCTGCTA	2750
	ATCGCCCTGA	TGCTTCAACT	AACATGTTGG	CTTGCGGGCG	TTCATGCTCA	2800
	GAAACAGGTT	TGGGACAAGC	ACTTCCAGGC	TAACACAGTC	AGAAATCGAA	2850
	ACGTACTCTC	AACAGTTCGC	TTAGGCATGG	AAGTTTTGCG	GCATTCGTGGC	2900
	TACACATAAA	CAAGGGAAGA	CTTACTCGTG	GCTGCAACCC	TACTAGCTCA	2950
15	AAATTTATTC	ACACATGGTT	ACGCTTTGGG	GAAATTATGA	TAATGATCCA	3000
	GATCACTTCT	GGCTAATAAA	AGATCAGAGC	TCTAGAGATC	TGTGTGTTGG	3050
	TTTTTTGTGG	ATCTGCTGTG	CCTTCTAGTT	GCCAGCCATC	TGTTGTTTGC	3100
	CCCTCCCCCG	TGCCTTCCTT	GACCCTGGAA	GGTGCCACTC	CCACTGTCCT	3150
	TTCCTAATAA	AATGAGGAAA	TTGCATCGCA	TTGTCTGAGT	AGGTGTCATT	3200
20	CTATTCTGGG	GGGTGGGGTG	GGGCAGCACA	GCAAGGGGGA	GGATTGGGAA	3250
	GACAATAGCA	GGCATGCTGG	GGATGCGGTG	GGCTCTATGG	GTACCTCTCT	3300
	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCGGTAC	CTCTCTCTCT	3350
	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CGGTACCAGG	TGCTGAAGAA	3400
	TTGACCCGGT	GACCAAAGGT	GCCTTTTATC	ATCACTTTAA	AAATAAAAAA	3450
25	CAATTACTCA	GTGCCTGTTA	TAAGCAGCAA	TTAATTATGA	TTGATGCCTA	3500
	CATCACAACA	AAAACGTATT	TAACAAATGG	TTGGTCTGCC	TTAGAAAAGTA	3550
	TATTTGAACA	TTATCTTGAT	TATATATTG	ATAATAATAA	AAACCTTATC	3600
	CCTATCCAAG	AAGTGATGCC	TATCATTTGG	TGGAATGAAC	TTGAAAAAAA	3650
	TTAGCCCTGA	ATACATTACT	GGTAAGGTAA	ACGCCATTGT	CAGCAAATTG	3700
30	ATCCAAGAGA	ACCAACTTAA	AGCTTTCCTG	ACGGAATGTT	AATTCCTCGT	3750
	GACCCGTAGC	ACTGATGAAT	CCCCTAATGA	TTTTGGTAAA	AATCATTAAG	3800
	TTAAGGTGGA	TACACATCTT	GTCATATGAT	CCCGGTAATG	TGAGTTAGCT	3850
	CACTCATTAG	GCACCCAGG	CTTTACACTT	TATGCTTCCG	GCTCGTATGT	3900
	TGTGTGGAAT	TGTGAGCGGA	TAACAATTTT	ACACAGGAAA	CAGCTATGAC	3950
35	CATGATPACG	CCAAGCGCGC	AATTAACCCCT	CACTAAAGGG	AACAAAAGCT	4000
	GGAGCTCCAC	CGCGGTGGCG	GCCGCTCTAG	AACTAGTGGG	TCCCCCGGGC	4050
	TGCAGGAATT	CGATATCAAG	CTTATCGATA	CCGCTGACCT	CGAGGGGGGG	4100
	CCCGGTACCC	AATTCGCCCT	ATAGTGAGTC	GTATTACGCG	CGCTCACTGG	4150
	CCGTCGTTTT	ACAACGTCGT	GACTGGGAAA	ACCCTGGCGT	TACCCAACCT	4200
40	AATCGCCTTG	CAGCACATCC	CCCTTTTCGCC	AGCTGGCGTA	ATAGCGAAGA	4250
	GGCCCGCACC	GATCGCCCTT	CCCAACAGTT	GCGCAGCCTG	AATGGCGAAT	4300
	GGAAATTGTA	AGCGTTAATA	TTTTGTAAAA	ATTCGCGTTA	AATTTTGTGT	4350
	AAATCAGCTC	ATTTTTTAAC	CAATAGGCCG	AAATCGGCAA	AATCCCTTAT	4400
	AAATCAAAAAG	AATAGACCGA	GATAGGGTTG	AGTGTGTGTC	CAGTTTGGA	4450
45	CAAGAGTCCA	CTATTAAAGA	ACGTGGACTC	CAACGTCAAA	GGGCGAAAAA	4500
	CCGTCTATCA	GGGCGATGGC	CCACIACTCC	GGGATCATAT	GACAAGATGT	4550
	GTATCCACCT	TAACTTAATG	ATTTTACCA	AAATCATTAG	GGGATTCATC	4600
	AGTGCTCAGG	GTCAACGAGA	ATTAACATTC	CGTCAGGAAA	GCTTATGATG	4650
	ATGATGTGCT	TAAAACTTA	CTCAATGGCT	GGTTATGCAT	ATCGCAATAC	4700
50	ATGCGAAAAA	CCTAAAAGAG	CTTGCCGATA	AAAAAGGCCA	ATTTATTGCT	4750
	ATTTACCGCG	GCTTTTATTT	GAGCTTGAAA	GATAAATAAA	ATAGATAGGT	4800
	TTTATTTGAA	GCTAAATCTT	CTTTATCGTA	AAAAATGCCC	TCTTGGGTTA	4850
	TCAAGAGGGT	CATTATATTT	CGCGGAATAA	CATCATTGTT	TGACCAAATA	4900
	ACTAAGCACT	TGICTCCTGT	TTACTCCCTT	GAGCTTGAGG	GGTTAACATG	4950
55	AAGGTCATCG	ATAGCAGGAT	AATAATACAG	TAAAACGCTA	AACCAATAAT	5000
	CCAAATCCAG	CCATCCCAAA	TTGGTAGTGA	ATGATTATAA	ATAACAGCAA	5050
	ACAGTAATGG	GCCAATAACA	CCGGTTGCAT	TGGTAAGGCT	CACCAATAAT	5100
	CCCTGTAAAG	CACCTTGCTG	ATGACTCTTT	GTTTGGATAG	ACATCACTCC	5150
	CTGTAATGCA	GGTAAAGCGA	TCCCACCACC	AGCCAATAAA	ATTAAACAG	5200
60	GGAAAACTAA	CCAACCTTCA	GATATAAACG	CTAAAAAGGC	AAATGCACTA	5250
	CTATCTGCAA	TAAATCCGAG	CAGTACTGCC	GTTTTTTCGC	CCATTTAGTG	5300

	GCTATTCTTC	CTGCCACAAA	GGCTTGGAAT	ACTGAGTGTA	AAAGACCAAG	5350
	ACCCGTAATG	AAAAGCCAAC	CATCATGCTA	TTATCATCA	CGATTTCTGT	5400
	AATAGCACCA	CACCGTGCTG	GATTGGCTAT	CAATGCGCTG	AAATAATAAT	5450
	CAACAAATGG	CATCGTTAAA	TAAGTGATGT	ATACCGATCA	GCTTTTGTTC	5500
5	CCTTTAGTGA	GGGTTAATTG	CGCGCTTGGC	GTAATCATGG	TCATAGCTGT	5550
	TTCCTGTGTG	AAATTGTTAT	CCGCTCACAA	TTCCACACAA	CATACGAGCC	5600
	GGAAGCATAA	AGTGTAAAGC	CTGGGGTGCC	TAATGAGTGA	GCTAACTCAC	5650
	ATTAATTGCG	TTGCGCTCAC	TGCCCCGCTT	CCAGTCGGGA	AACCTGTCGT	5700
	GCCAGCTGCA	TTAATGAATC	GGCCAACGCG	CGGGGAGAGG	CGGTTTGCCT	5750
10	ATTGGGCGCT	CTTCCGCTTC	CTCGCTCACT	GACTCGCTGC	GCTCGGTCGT	5800
	TGGGCTGCGG	CGAGCGGTAT	CAGCTCACTC	AAAGGCGGTA	ATACGGTTAT	5850
	CCACAGAATC	AGGGGATAAC	GCAGGAAAGA	ACATGTGAGC	AAAAGGCCAG	5900
	CAAAAGGCCA	GGAAACGTAA	AAAGGCCGCG	TTGCTGGCGT	TTTCCATAG	5950
	GCTCCGCCCC	CCTGACGAGC	ATCACAAAAA	TCGACGCTCA	AGTCAGAGGT	6000
15	GGCGAAACCC	GACAGGACTA	TAAAGATACC	AGGCGTTTCC	CCCTGGAAGC	6050
	TCCCTCGTGC	GCTCTCCTGT	TCCGACCCTG	CCGCTTACCG	GATACCTGTC	6100
	CCGCTTTTCTC	CCTTCGGGAA	CGGTGGCGCT	TTCTCATAGC	TCACGCTGTA	6150
	GGTATCTCAG	TTGCGGTAG	GTCGTTGCGT	CCAAGCTGGG	CTGTGTGCAC	6200
	GAACCCCCCG	TTGAGCCCGA	CCGCTGCGCC	TTATCCGGTA	ACTATCGTCT	6250
20	TGAGTCCAAC	CCGGTAAGAC	ACGACTTATC	GCCACTGGCA	GCAGCCACTG	6300
	GTAACAGGAT	TAGCAGAGCG	AGGTATGTAG	GCGGTGCTAC	AGAGTTCCTG	6350
	AAGTGGTGGC	CTAATACGG	CTACACTAGA	AGGACAGTAT	TTGGTATCTG	6400
	CGCTCTGCTG	AAGCCAGTTA	CCTTCGGAAA	AAGAGTTGGT	AGCTCTTGAT	6450
	CCGGCAAACA	AACCACCGCT	GGTAGCGGTG	GTTTTTTTGT	TTGCAAGCAG	6500
25	CAGATTACGC	GCAGAAAAAA	AGGATCTCAA	GAAGATCCTT	TGATCTTTTC	6550
	TACGGGGTCT	GACGCTCAGT	GGAACGAAAA	CTCAGTTAA	GGGATTTTGG	6600
	TCATGAGATT	ATCAAAAAGG	ATCTTCACCT	AGATCCTTTT	AAATTAAAAA	6650
	TGAAGTTTTA	AATCAATCTA	AAGTATATAT	GAGTAACTT	GGTCTGACAG	6700
	TTACCAATGC	TTAATCAGTG	AGGCACCTAT	CTCAGCGATC	TGTCTATTTC	6750
30	GTTCATCCAT	AGTTGCCTGA	CTCCCCGTCG	TGTAGATAAC	TACGATACGG	6800
	GAGGGCTTAC	CATCTGGCCC	CAGTGCTGCA	ATGATACCGC	GAGACCCACG	6850
	CTCACCGGCT	CCAGATTAT	CAGCAATAAA	CCAGCCAGCC	GGAAGGGCCG	6900
	AGCGCAGAAG	TGGTCCTGCA	ACTTTATCCG	CCTCCATCCA	GTCTATTAAT	6950
	TGTTGCCCGG	AAGCTAGAGT	AAGTAGTTCG	CCAGTTAATA	GTTTGCGCAA	7000
35	CGTTGTTGCC	ATTGCTACAG	GCATCGTGGT	GTCACGCTCG	TCGTTTGGTA	7050
	TGGCTTCATT	CAGCTCCGGT	TCCCAACGAT	CAAGGCGAGT	TACATGATCC	7100
	CCCATGTTGT	GCAAAAAAGC	GGTAGCTCC	TTGCGTCCTC	CGATCGTTGT	7150
	CAGAAGTAAG	TTGGCCGCG	TGTTATCACT	CATGTTTATG	GCAGCACTGC	7200
	ATAATTCTCT	TACTGTCAATG	CCATCCGTAA	GATGCTTTTC	TGTGACTGGT	7250
40	GAGTACTCAA	CCAAGTCATT	CTGAGAATAG	TGTATGCGGC	GACCGAGTTG	7300
	CTCTTGCCCG	GCGTCAATAC	GGGATAATAC	CGCGCCACAT	AGCAGAACTT	7350
	TAAAAGTGCT	CATCATTTGA	AAACGTTCTT	CGGGGCGAAA	ACTCTCAAGG	7400
	ATCTTACCGC	TGTTGAGATC	CAGTTCGATG	TAACCCACTC	GTGCACCCAA	7450
	CTGATCTTCA	GCATCTTTTA	CTTTCACCAG	CGTTTCTGGG	TGAGCAAAAA	7500
45	CAGGAAGGCA	AAATGCCGCA	AAAAAGGGAA	TAAGGGCGAC	ACGGAAATGT	7550
	TGAATACTCA	TACTCTTCCT	TTTTCAATAT	TATTGAAGCA	TTTATCAGGG	7600
	TTATTGICTC	ATGAGCGGAT	ACATATTTGA	ATGTATTTAG	AAAAATAAAC	7650
	AAATAGGGGT	TCCGCGCACA	TTTCCCCGAA	AAGTGCCAC		7689
50	SEQ ID NO:8 (modified Kozak sequence)					
	ACCATG					
	SEQ ID NO:9 (a Kozak sequence)					
	ACCATGG					
55	SEQ ID NO:10 (a Kozak sequence)					
	ACCATGT					
	SEQ ID NO:11 (a Kozak sequence)					
60	AAGATGT					

[illegible]

TCCCTCGAAC CATGAACACT CCTCCAGCTG AATTTCACAA TTCCTCTGTC  
 ATCTGCCAGG CCATTAAGTT ATTCATGGAA GATCTTTGAG GAACACTGCA  
 AGTTCATATC ATAAACACAT TTGAAATTGA GTATTGTTTT GCATTGTATG  
 GAGCTATGTT TTGCTGTATC CTCAGAAAAA AAGTTTGTTA TAAAGCATT  
 5 ACACCCATAA AAAGATAGAT TTAAATATTC CAGCTATAGG AAAGAAAGTG  
 CGTCTGCTCT TCACTCTAGT CTCAGTTGGC TCCTTCACAT GCATGCTTCT  
 TTATTTCTCC TATTTTGTCA AGAAAATAAT AGGTCACGTC TTGTTCTCAC  
 TTATGTCTTG CCTAGCATGG CTCAGATGCA CGTTGTAGAT ACAAGAAGGA  
 10 TCAAATGAAA CAGACTTCTG GTCTGTTACT ACAACCATAG TAATAAGCAC  
 ACTAACTAAT AATTGCTAAT TATGTTTTCC ATCTCTAAGG TTCCACATT  
 TTTCTGTTTT CTAAAGATC CCATTATCTG GTTGTAACG AAGCTCAATG  
 GAACATGAGC AATATTTCCC AGTCTTCTCT CCCATCCAAC AGTCCTGATG  
 GATTAGCAGA ACAGGCAGAA AACACATTGT TACCCAGAAT TAAAACTAA  
 TATTTGCTCT CCATTCAATC CAAAATGGAC CTATTGAAAC TAAAATCTAA  
 15 CCCAATCCCA TTAAATGATT TCTATGGCGT CAAAGGTCAG ACTTCTGAAG  
 GGAACCTGTG GGTGGGTCAC AATTCAGGCT ATATATTCCC CAGGGCTCAG

SEQ ID NO:22 (chicken ovalbumin ehancer)  
 cggggctgca gaaaaatgcc aggtggacta tgaactcaca tccaaaggag  
 20 cttgacctga tacctgattt tottcaaaact ggggaaacaa cacaatccca caaaacagct  
 cagagagaaa ccatcactga tggctacagc accaaggtat gcaatggcaa tccattcgac  
 attcatctgt gacctgagca aaatgattta tctctccatg aatggttgct tctttccctc  
 atgaaaaggc aatttccaca ctccacaatat gcaacaaaga caaacagaga acaattaatg  
 tgctccttcc taatgtcaaa attgtagtgg caaagaggag acaaaaatct caagttctga  
 25 gtagggtttta gtgattggat aagaggcttt gacctgtgag ctccacctgga ctccatattcc  
 ttttgataa aaagtgtctt tataactttc aggtctccga gtctttattc atgagactgt  
 tgggttaggg acagaccac aatgaaatgc ctggcatagg aaagggcagc agagccttag  
 ctgacctttt cttgggacaa gcattgtcaa acaatgtgtg acaaaactat ttgtactgct  
 ttgcacagct gtgctgggca gggcaatcca ttgccaccta tcccaggtaa ccttccaact  
 30 gcaagaagat tgttgcttac tctctctaga

SEQ ID NO:23 (5' untranslated region)  
 GTGGATCAACATACAGCTAGAAAGCTGTATTGCCTTTAGCACTCAAGCTCAAAGACAAC TCAGAGTTTC  
 ACC  
 35

SEQ ID NO:24 (putative cap site)  
 ACATACAGCTAG AAAGCTGTAT TGCCTTTAGC ACTCAAGCTC AAAAGACAAC TCAGAGTTCA

SEQ ID NO:25 (Chicken Ovalbumin Signal Sequence)  
 40 ATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTGTATG TATTCAAGGA GCTCAAAGTC  
 CACCATGCCA ATGAGAACAT CTTCTACTGC CCCATTGCCA TCATGTCAGC TCTAGCCATG  
 GTATACCTGG GTGCAAAAGA CAGCACCAGG ACACAGATAA ATAAGGTTGT TCGCTTTGAT  
 AAAC TTCCAG GATTCGGAGA CAGTATTGAA GCTCAGTGTG GCACATCTGT AAACGTTTAC  
 TCTTCACTTA GAGACATCCT CAACCAAAATC ACCAAACCAA ATGATGTTTA TTCGTTTACG  
 45 CTTGCCAGTA GACTTTATGC TGAAGAGAGA TACCAATCC TGCCAGAATA CTGCAAGTGT  
 GTGAAGGAAC TGTATAGAGG AGGCTTGGAA CCTATCAACT TTCAAACAGC TGCAGATCAA  
 GCCAGAGAGC TCATCAATTC CTGGGTAGAA AGTCAGACAA ATGGAATTAT CAGAAATGTC  
 CTTCAAGCAA GCTCCGTGGA TTCTCAAAC GCAATGGTTC TGGTTAATGC CATTGTCTTC  
 AAAGGACTGT GGGAGAAAAC ATTTAAGGAT GAAGACACAC AAGCAATGCC TTTCAGAGTG  
 50 ACTGAGCAAG AAAGCAAACC TGTGCAGATG ATGTACCAGA TTGGTTTATT TAGAGTGGCA  
 TCAATGGCTT CTGAGAAAAT GAAGATCCTG GAGCTTCCAT TTGCCAGTGG GACAATGAGC  
 ATGTTGGTGC TGTGCTTGA TGAAGTCTCA GGCCTTGAGC AGCTTGAGAG TATAATCAAC  
 TTTGAAAAAC TGACTGAATG GACCAGTTCT AATGTTATGG AAGAGAGGAA GATCAAAGTG  
 TACTTACCTC GCATGAAGAT GGAGGAAAAA TACAACCTCA CATCTGTCTT AATGGCTATG  
 55 GGCATTACTG ACGTGTTTAG CTCTTCAGCC AATCTGTCTG GCATCTCCTC AGCAGAGAGC  
 CTGAAGATAT CTCAAGCTGT CCATGCAGCA CATGCAGAAA TCAATGAAGC AGGCAGAGAG  
 GTGGTAGGGT CAGCAGAGGC TGGAGTGGAT GCTGCAAGCG TCTCTGAAGA ATTTAGGGCT  
 GACCATCCAT TCCTCTTCTG TATCAAGCAC ATCGCAACCA ACGCCGTTCT CTCTTTTGGC  
 60 AGATGTGTTT CCCCT

SEQ ID NO:26 (Chicken Ovalbumin Signal Sequence - shortened 50bp)

- ATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTTGATG TATTCAAGGA
- SEQ ID NO:27 (Chicken Ovalbumin Signal Sequence - shortened 100bp)  
 5 ATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTTGATG TATTCAAGGA GCTCAAAGTC  
 CACCATGCCA ATGAGAACAT CTTCTACTGC CCCATTGCCA
- SEQ ID NO:28 (vitellogenin targeting sequence)  
 10 ATGAGGGGGATCATACTGGCATTAGTGCTCACCTTGTAGGCAGCCAGAAGTTTGACATTGGT
- SEQ ID NO:29 (pro-insulin sequence)  
 TTGTGAACCAACACCTGTGCGGCTCACACCTGGTGGAGCTCTCTACCTAGTGTGCGGGGAACGAGGC  
 TTCTTCTACACACCCAGACCCGCCGGGAGGCAGAGGACCTGCAGGTGGGGCAGGTGGAGCTGGGCGGG  
 15 GGCCCTGGTGCAGGCAGCCTGCAGCCCTTGGCCCTGGAGGGGTCCCTGCAGAGCGTGGCATTGTGGAA  
 CAATGCTGTACCAGCATCTGCTCCCTCTACCAGCTGGGAACTCTGCAACTAG
- SEQ ID NO:30 (p146 protein)  
 KYKKALKKLAKLL
- 20 SEQ ID NO:31 (p146 coding sequence)  
 AAATACAAAAAAGCACTGAAAAAAGTGGCAAACTGCTG
- SEQ ID NO:32 (spacer)  
 25 (GPGG)<sub>x</sub>
- SEQ ID NO:33 (spacer)  
 GPGGGPGGPGG
- 30 SEQ ID NO:34 (spacer)  
 GGGGSGGGSGGGGS
- SEQ ID NO:35 (spacer)  
 35 GGGGSGGGSGGGSGGGGS
- SEQ ID NO:36 (repeat domain in TAG spacer sequence)  
 Pro Ala Asp Asp Ala
- SEQ ID NO:37 (TAG spacer sequence)  
 40 Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala  
 Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp
- SEQ ID NO:38 (gp41 epitope)  
 45 Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys Gly Trp Ile Gly Leu Leu
- SEQ ID NO:39 (polynucleotide sequence encoding gp41 epitope)  
 Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly  
 Ser Cys Gly Trp Ile Gly Leu Leu Asp Asp Asp Asp Lys
- 50 SEQ ID NO:40 (enterokinase cleavage site)  
 DDDDK
- SEQ ID NO:41 (TAG sequence)  
 Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala  
 55 Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile  
 Leu Lys Gly Ser Cys Gly Trp Ile Gly Leu Leu Asp Asp Asp Asp Lys
- SEQ ID NO:42 (altered transposase Hef forward primer)  
 ATCTCGAGACCATGTGTGAAGTGTGATATTTACATGATTCTCTTACC

SEQ ID NO:43 (altered transposase Her reverse primer)  
GATTGATCATTATCATAATTTCCCAAGCGTAACC

5

SEQ ID NO: 44 GnRH:Phor 11  
Met-Glu-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly-Lys-Phe-Ala-Ile-Cys-Lys-  
Lys-Phe-Ala-Ile-Cys-OCH

10

SEQ ID NO: 45 GNRH/Phor14  
EHWSYGLRPGKFAKFAKFAKFAK

15

SEQ ID NO: 46 Phor14::Beta-LH Sequence  
MKFAKFAKFAKFAKSYAVALSCQCALCRR

20

SEQ ID NO:47 (pTnMCS (CMV-prepro-HCPro-ProLys-LC-CPA))

```

1  ctgacgcgcc ctgtagcgcc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga
61  cgcgtacact tgcacgcgcc ctacgcgccg ctcccttctgc tttcttccct tcctttctcg
121  ccacgtttcg cggcatcaga ttggtatttg gccattgcat acgttgatgc catatcataa
181  tatgtacatt tatattgggt catgtccaac attaccgcca tgttgacatt gattattgac
241  tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg
301  cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc ccgcccatt
361  gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca
421  atgggtggag tatttacggt aaactgcccc ctgggcagta catcaagtgt atcatatgcc
481  aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta
541  catgacctta tgggactttc ctacttgcca gtacatctac gtatttagtca tcgctattac
601  catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg
661  atttccaagt ctccaccoca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg
721  ggactttcca aaatgtcgta acaactccgc ccattgacg caaatggcg gtaggcgtgt
781  acggtgggag gtctatataa gcagagctcg ttagtgaac cgtcagatcg cctggagacg
841  ccattccacg tgttttgacc tccatagaag acaccgggac cgtccagcc tccgcgccg
901  ggaacgggtg attggaacgc ggattcccg tgccaagagt gacgtaagta ccgctatag
961  actctatagg cacacccctt tggctcttat gcattgctata ctgttttttg cttggggcct
1021  atacaccccc gcttccttat gctatagggt atggtatagc ttgacctata ggtgtgggtt
1081  attgaccatt attgaccatt cccctatttg tgacgatact ttccattact aatccataac
1141  atggctcttt gccacaacta tctctatttg ctatatgcca atactctgtc cttcagagac
1201  tgacacggac tctgtatttt tacaggatgg ggtcccatat attatttaca aattccata
1261  tacacaacac cgttccccgc tgcccgcagt ttttattaaa catagcgttg gatctccacg
1321  cgaatctcgg gtaactgttc cggacatggg ctcttctccg gtacggcgcg agcttccaca
1381  tccgagccct ggtcccattg ctccagcgcc tcatggtcgc tcggcagctc cttgctccta
1441  acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag
1501  gccgtggcgg taggggtatg gtctgaaaat gagcgtggag attgggctcg cagggtgac
1561  gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgtatc
1621  tgataagagt cagaggtaac tccggttgcg gtgctgttaa cgtggaggcg cagtgtagt
1681  tgagcagtag tctgtgtcgc cgcgcgcgcc accagacata atagctgaca gactaacaga
1741  ctgttccctt ccatgggtct tttctgcagt caccgtcgga ccatgtgcga actcgatatt
1801  ttacacgact ctctttacca attctgcccc gaattacact taaaacgact caacagctta
1861  acgttggtct gccacgcatt acttgactgt aaaactctca ctcttaccga acttgccgt
1921  aacctgccaa ccaaaagcag aacaaaacat aacatcaaac gaatcgaccg attgttaggt
1981  aatcgtcacc tccacaaaga gcgactcgct gtataccggt ggcatgctag ctttatctgt
2041  tcgggcaata cgtatgccat tgtacttgtt gactgtgtct atattcgtga gcaaaaacga
2101  cttatggtat tgcgagcttc agtcgcacta caggtctggt ctgttactct ttatgagaaa
2161  gcgttccgcg tttcagagca atgttcaaag aaagctcatg accaatttct agccgacctt
2221  gcgagcatc taaccagtaa caccacaccg ctcatgttca gtgatgctgg ctttaaagt
2281  ccattggtata aatccgttga gaagctgggt tggtagctgt taagtcgagt aagaggaaaa
2341  gtacaatatg cagacctagg agcggaaaaa tggaaaacta tcagcaactt acatgatatg
2401  tcatctagtc actcaaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca
2461  tgccaaattc tattgtataa atctcgctct aaaggccgaa aaaatcagcg ctgcacacgg
2521  actcattgtc accaccgctc acctaaaatc tactcagcgt cggcaaaagg gccatgggtt
2581  ctagcaacta acttacctgt tgaattcga acacccaaac aacttgtaa tatctattcg
2641  aagcgaatgc agattgaaga aacctccga gacttgaaaa gtcctgcta cggactaggc
2701  ctacgccata gcgaacgag cagctcagag cgttttgata tcatgctgct aatcgccctg
2761  atgcttcaac taacatgttg gcttgcgggc gttcatgctc aaaaacaagg ttgggacaag
2821  cacttccagg ctaacacagt cagaatcga aacgtactct caacagttcg cttaggcatg
2881  gaagttttgc ggcattctcg ctacacaata acaagggaag acttactcgt ggctgcaacc

```

50

60

65

	2941	ctactagctc	aaaattttatt	cacacatggt	tacgcttttg	ggaaattatg	aggggatcgc
	3001	tctagagcga	tccgggatct	cgggaaaagc	gttgggtgacc	aaaggtgcct	tttatcatca
	3061	ctttaaaaat	aaaaaacaat	tactcagtg	ctgttataag	cagcaattaa	ttatgattga
5	3121	tgcctacatc	acaacaaaaa	ctgattttaac	aaatgggttg	tctgccttag	aaagtatat
	3181	tgaacattat	cttgattata	ttattgataa	taataaaaaa	cttatcccta	tccaagaagt
	3241	gatgcctatc	attgggttga	atgaacttga	aaaaaattag	ccttgaatac	attactggta
	3301	aggtaaacgc	cattgtcagc	aaattgatcc	aagagaacca	acttaaaagt	ttcctgacgg
	3361	aatgttaatt	ctcgttgacc	ctgagcactg	atgaatcccc	taatgatttt	ggtaaaaaatc
10	3421	attaagttaa	ggtggatata	catcttgtca	tatgatcccg	gtaatgtgag	ttagctcact
	3481	cattagggcac	cccaggcttt	acactttatg	cttccggctc	gtatgtttgtg	tggaattgtg
	3541	agcggataac	aatttcacac	aggaacacgc	tatgaccatg	attacgcaa	gcgcgcaatt
	3601	aacctcact	aaagggaaac	aaagctggag	ctccaccgcg	gtggcggccg	ctctagaact
	3661	agtggatccc	cggggctgca	ggaattcgat	atcaagctta	togataccgc	tgacctcgag
15	3721	cattaggttg	gctattggcc	attgcatacg	ttgtatccat	atcataatat	gtcaatttat
	3781	attggctcat	gtccaacatt	accgccatgt	tgacattgat	tattgactag	ttattaatag
	3841	taatcaatta	cggggctcatt	agttcatagc	ccatatatgg	agttcccggt	tacataactt
	3901	acgttaaatg	gccgcgctgg	ctgacccccc	aacgaccccc	gccatttgac	gtcaataatg
	3961	acgtatgttc	ccatagtaac	gccaataggg	actttccatt	gacgtcaatg	ggtggagtat
20	4021	ttacggtaaa	ctgcccactt	ggcagtagat	caagtgtatc	atatgtcaag	tacgccccct
	4081	attgacgtca	atgacggtaa	atggcccgc	tggcattatg	cccagtagat	gaccttatgg
	4141	cgctttccta	cttggcagta	catctacgta	ttagtcatcg	ctattaccat	ggtgatcgcg
	4201	ttttggcagt	acatcaatgg	gggtggatag	cggtttgact	cacggggatt	tccaagtctt
	4261	cacccatttg	acgtcaatgg	gagtttgttt	tggcaccaaa	atcaacggga	ctttccaaaa
25	4321	tgtcgtaaca	actccgcccc	attgacgcaa	atgggcggta	ggcgtgtacg	gtgggaggtc
	4381	tataaagca	gagctcgttt	agtgaaacct	cagatcgcc	ggagacgcca	tccacgtgt
	4441	tttgacctcc	atagaagaca	ccgggaccca	tccagcctcc	gcggccggga	acggtgcatt
	4501	ggaacgggga	ttccccgtgc	caagagtgc	gtaagtaccg	cctatagact	ctataggcac
	4561	acccctttgg	ctcttatgca	tgctatactg	tttttgcttt	ggggcctata	cacccccgat
30	4621	tccttatgct	ataggtgatg	gtatagctta	gcctataggt	gtgggttatt	gaccattatt
	4681	gacctctccc	ctattgttga	cgatactttc	cattactaat	ccataacatg	gctctttgcc
	4741	acaactatct	ctattggcta	tatgccaaata	ctctgtcctt	cagagactga	cacggactct
	4801	gtatttttac	aggatggggg	cccattttat	atttacaat	tcacatatac	aaacaacg
	4861	tccccgtgc	ccgcagtttt	tattaaacat	agcgtgggat	ctccacgcga	atctcgggta
	4921	cgtgttccgg	acatgggtcc	ttctccggta	gcggcgggag	ttccacatcc	gagccctgg
35	4981	ccatgcctc	cagcggctca	tggtcgctcg	gcagctcctt	gctcctaaca	gtggagccca
	5041	gacttaggca	cagcacatg	cccaccacca	ccagtgtgcc	gcacaaggcc	gtggcggtag
	5101	ggtatgtgtc	tgaanaatgag	cgtggagatt	gggctcgcac	ggctgacgca	gatggaagac
	5161	ttaaggcagc	ggcagaagaa	gatgcaggca	gctgagttgt	tgtattctga	taagagttag
40	5221	aggtaactcc	cgttgccgtg	ctgttaacgg	tggaggcgag	tgtagtctga	gcagtagctg
	5281	ttgctgccgc	gcgcgccacc	agacataata	gctgacagac	taacagactg	ttcctttcca
	5341	tgggtctttt	ctgcagtcac	cgtcggatca	atcattcatc	togtgacttc	ttcgtgtgtg
	5401	gtgtttacct	atatatctaa	atttaatat	togtttatta	aaatttaata	tatttcgacg
	5461	atgaatttct	caaggatatt	ttctctcgtg	ttcgtcttgg	ttctggcttt	gtcaacagtt
45	5521	toggctgcgc	cagagccgaa	aggtaccacg	gtgcagctgc	aggagtccgg	gggaggcttg
	5581	gtaaaagccg	gggggtccct	tagagtctcc	tgtgcagcct	ctggattcac	tttcagaaac
	5641	ccctgggatg	gctgggtccg	ccaggctcca	gggaaggggc	tggagtgggt	cggccgtatt
	5701	aaaagcaaaa	ttgatgttgg	gacaacagac	tatgctgcac	ccgtgaaagg	cagattcacc
	5761	atctcaagag	atgattcaaa	aaacacgtta	tatctgcaaa	tgaatagcct	gaaagccgag
50	5821	gacacagccg	tatattactg	taccacgggg	attatgataa	catttggggg	agttatccct
	5881	cccccgaaat	ggggccaggg	aaccttggtc	accgtctcct	cagcctccac	caagggccca
	5941	tccgtcttcc	ccctggcacc	ctcctccaag	agcacctctg	ggggcacagc	ggcctggggc
	6001	tgccctggtc	aggactactt	ccccgaaccg	gtgacgggtg	cgtggaaact	aggcgccctg
	6061	accagcggcg	tgacacacct	tccggctgtc	ctacagtcct	caggactcta	cttcccttagc
55	6121	aacgtgggtg	cctgtccctc	cagcagcttg	ggcaccacga	cctacatctg	caacgtgaat
	6181	cacaagccca	gcaacaccaa	ggtggacaag	aaagttgagc	ccaaatcttg	tgacaaaact
	6241	cacacatgcc	caccgtgccc	agcacctgaa	ctcctggggg	gacogtcagt	cttccctctc
	6301	cccccaaaac	ccaaggacac	cctcatgac	tcccggaccc	ctgaggtcac	atgcgtgggtg
	6361	gtggacgtga	gccacgaaga	ccctgaggtc	aagttcaact	ggtacgtgga	cggcgtggag
60	6421	gtgcataatg	ccaagacaaa	gcccggggag	gagcagtaga	acagcacgta	ccgtgtgggtc
	6481	agcgtctctc	cgtctctgca	ccaggactgg	ctgaatggca	aggagtaca	gtgcaagggtc
	6541	tccaacaaaag	ccctcccgag	cccatcgag	aaaaccatct	ccaaagccaa	agggcagccc
	6601	cgagaaccac	aggtgtacac	cctgccccca	tcccgggatg	agctgaccaa	gaaccagggtc
	6661	agcctgacct	gcctgtgtca	aggcttctat	cccagcgaca	tgcctgtgga	gtgggagagc
65	6721	aatgggcagc	cggagaacaa	ctacaagacc	acgcctcccg	tgctggactc	cgacggctcc
	6781	ttcttctctc	acagcaagct	caccgtggac	aagagcaggt	ggcagcaggg	gaacgtcttc
	6841	tcatgctccg	tgatgcata	ggctctgcac	aaccactaca	cgcagaagag	cctctccctg
	6901	tctccgggta	aagcgccaga	gccgaaaaag	cttctctatg	agctgacaca	gccacccctg
	6961	gtgtcagtg	ccccaggaca	aacggccagg	atcacctgct	ctggagatgc	attgcccagaa



	7021	aaatatgttt	attggtacca	gcagaagtea	ggccaggccc	ctgtggtggt	catctatgag
	7081	gacagcaaac	gacctctcgg	gatccctgag	agattctctg	gctccagctc	agggacaatg
	7141	gccaccttga	ctatcagtg	ggcccagggtg	gaagatgaag	gtgactacta	ctgttactca
5	7201	actgacagca	gtgggttatca	tagggagggtg	ttcagcggag	ggaccaagct	gacctctcta
	7261	ggtcagccca	aggctgcccc	ctcggtcact	ctgttcccac	cctcctctga	ggagcttcaa
	7321	gccacaagg	ccacactggt	gtgtctcata	agtgaactcct	accggggagc	cgtgacagt
	7381	gcctggaagg	cagatagcag	ccccgtcaag	gcgggagtg	agaccaccac	accctccaaa
	7441	caaagcaaca	acaagtacgc	ggccagcagc	tacctgagcc	tgacgcttga	gcagtggaa
	7501	tcccacaaaa	gctacagctg	ccaggctcag	catgaaggga	gcaccgtgga	gaagacagt
10	7561	gccoctgcag	aatgttccac	gcggagggtg	ggaagggcc	tttttgaagg	gggaggaaac
	7621	ttcgcgccat	gactcctctc	gtgcccccg	cacggaacac	tgatgtgcag	agggccctct
	7681	gccattgtcg	cttcctctgc	ccttcctcgt	cactctgaat	gtggcttctt	tgctactgcc
	7741	acagcaagaa	ataaaatctc	aacatctaaa	tggttttctt	gagatttttc	aagagtcgtt
	7801	aagcacattc	cttccccagc	acccttgct	gcaggccagt	gccaggcacc	aacttggcta
15	7861	ctgctgcccc	tgagagaaat	ccagttcaat	attttccaaa	gcaaaatgga	ttacatatgc
	7921	cctagatcct	gattaacagg	tggtttgtat	tatctgtgct	ttcgcttcac	ccacattatc
	7981	ccattgcctc	ccctcgaggg	ggggcccggt	acccaattcg	ccctatagtg	agtcgtatta
	8041	cgcgcgctca	ctggcgctcg	ttttacaacg	tcgtgactgg	gaaaaccctg	gcgttaccga
	8101	acttaatcgc	cttgacgac	atcccccttt	cgccagctgg	cgtaatagcg	aagaggcccc
20	8161	caccgatcgc	ccttcccaac	agttgcgcag	cctgaatggc	gaatggaaat	tgtaagcgtt
	8221	aattatttgt	taaaattcgc	gttaaatctt	tggttaaatca	gtcatttttt	taaccaatag
	8281	gccgaaatcg	gcaaaatccc	ttataaatca	aaagaataga	ccgagatagg	gttgagtgtt
	8341	gttccagttt	ggaacaagag	tcactattta	aagaacgtgg	actccaacgt	caaaggcgca
	8401	aaaaccgtct	atcaggggcga	tggtccacta	ctccgggagc	atatgacaag	atgtgtatcc
25	8461	acottaaactt	aatgattttt	accaaaatca	ttagggggatt	catcagtgct	cagggtcaac
	8521	gagaattaac	attccgtcag	gaaagcttat	gatgatgatg	tgcttaaaaa	cttactcaat
	8581	ggctgggttat	gcataatcga	atacatcgca	aaaacctaaa	agagcttgcc	gataaaaaag
	8641	gccaatttat	tgctattttac	cgcggtcttt	tattgagctt	gaaagataaa	taaaatagat
	8701	aggtttttat	tgaagctaaa	tcctctttat	cgtaaaaaat	gccctcttgg	gttatcaaga
30	8761	gggtcattat	atttcgcgga	ataacatcat	ttggtgacga	aataactaag	cacttgtctc
	8821	ctgtttactc	ccctgagctt	gaggggttaa	catgaaggtc	atcgatagca	ggataataat
	8881	acagtaaaac	gctaaaccaa	taatccaaat	ccagccatcc	caaatggta	gtgaatgatt
	8941	ataaataaca	gcaaacagta	atgggccaat	aacaccggtt	gcatttggtta	ggctcaccaa
	9001	taatccctgt	aaagcacctt	gctgatgact	ctttgttttg	atagacatca	ctccctgtaa
35	9061	tgccaggtaaa	gcgatccca	caccagccaa	taaaattaaa	acagggaata	ctaaccaacc
	9121	ttcagatata	aacgctaaaa	agcctaaatgc	actactatct	gcaataaatc	cgagcagtac
	9181	tgccgttttt	tcgcccattt	agtggctatt	cttcctgcca	caaaggcttg	gaatactgag
	9241	tgtaaaagac	caagaccctg	aatgaaaagc	caaccatcat	gctatttcac	atcacgattt
40	9301	ctgtaatagc	accacaccgt	gctggatttg	ctatcaatgc	gctgaaataa	taatacaaca
	9361	atggcatcgt	taaataagtg	atgtataccg	atcagctttt	gttcccttta	gtgaggggta
	9421	attgcgcgct	tggtgtaatc	atggtcatag	ctgtttcctg	tgtagaaattg	ttatccgctc
	9481	acaattccac	acaacatacg	agccgggaagc	ataaagtgtg	aagcctgggg	tgccaatatga
	9541	gtgagctaac	tcacattaat	tgcggttcgc	tcactgccc	ctttccagtc	gggaaacctg
45	9601	tcgtgccagc	tgcatataatg	aatcgcccaa	cgccggggga	gaggcggttt	gcgtattggg
	9661	cgctcttcog	cttcctcgc	cactgactcg	ctgcgctcgg	tcgttcggct	gcggcgagcg
	9721	gtatcagctc	actcaaagcg	ggtaatacgg	ttatccacag	aatcaggggga	taacgcagga
	9781	aagaacatgt	gagcaaaaag	ccagcaaaaag	gccagggaacc	gtaaaaaagc	cgcgttgctg
	9841	gcgtttttcc	ataggctccg	ccccctgac	gagcatcaca	aaaatcgacg	ctcaagtcag
	9901	aggtggcgaa	accgacagcg	actataaaga	taccaggcgt	ttccccctgg	aagctccctc
50	9961	gtgcgctctc	ctgttccgac	cctgcgcgtt	accggatacc	tgctccgctt	tcctccctcg
	10021	ggaagcgtgg	cgctttctca	tagctcacgc	tgtaggtatc	tcagttcggg	gtaggtcgtt
	10081	cgctccaagc	tggtctgtgt	gcacgaaccc	ccgtttcagc	ccgaccgctg	cgcttatcc
	10141	ggtaactatc	gtcttgagtc	caaccgggta	agacacgact	tatcgccact	ggcagcagcc
55	10201	actggtaaca	ggattagcag	agcgagggtat	gtaggcggtg	ctacagagtt	cttgaagtgg
	10261	tggtcctaact	acggctacac	tagaaggaca	gtatttggtg	tcgtcgctct	gctgaagcca
	10321	gttaccttcg	gaaaaagagt	tggtagctct	tgatccggca	aaacaaaccac	cgctggtagc
	10381	ggtgtgtttt	ttgtttgcaa	gcagcagatt	acgcgcagaa	aaaaaggatc	tcaagaagat
	10441	cctttgatct	tttctacggg	gtctgacgct	cagtggaaacg	aaaactcacg	tttaaggatt
60	10501	ttgggtcatga	gattatcaaa	aaggatcttc	acctagatcc	ttttaaatta	aaaatgaagt
	10561	tttaaatcaa	tctaaagtat	atatgagtaa	acttggtctg	acagttacca	atgcttaatc
	10621	agtggagcac	ctatctcagc	gatctgtcta	tttcgttcat	ccatagttgc	ctgactcccc
	10681	gtcgtgtaga	taactacgat	acgggagggc	ttaccatctg	gccccagtcg	tgcaatgata
	10741	ccgcgagacc	cacgctcacc	ggctccagat	ttatcagcaa	taaacagacc	agccggaaag
65	10801	gccgagcgca	gaagtggctc	tgcaacttta	tcgcctcca	tcagctctat	taattgttgc
	10861	cgggaagcta	gagtaagtag	tgcccgagtt	aatagtttgc	gcaacgttgt	tgccattgct
	10921	acaggcatcg	tggtgtcacg	ctcgtcgttt	ggtaggctt	cattcagctc	cggttcccaa
	10981	cgatcaaggc	gagttacatg	atcccccatg	ttgtgcaaaa	aagcggttag	ctccttcggt
	11041	cctccgatcg	ttgtcagaag	taagttggcc	gcagtgttat	cactcatggt	tatggcagca

11101 ctgcataatt ctcttactgt catgccatcc gtaagatgct tttctgtgac tgggtgagtac  
 11161 tcaaccaagt cttcttgaga atagtgtatg cggcgaccga gttgctcttg cccggcgctca  
 11221 atacgggata ataccgcgcc acatagcaga actttaaaag tgctcatcat tggaaaacgt  
 5 11281 tcttcggggc gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaaccc  
 11341 actcgtgcac ccaactgato ttacgcatct ttactttca ccagcgtttc tgggtgagca  
 11401 aaaaacaggaa ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa atgttgata  
 11461 ctcatactct tcctttttca atattattga agcatttata agggttattg tctcatgagc  
 11521 ggatacatat ttgaatgtat ttagaaaaat aaacaaatag ggggtccgcg cacatttccc  
 11581 cgaaaagtgc cac  
 10  
 SEQ ID NO: 48 (pTnMCS (CMV-prepro-HCPro-CPA) )  
 1 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga  
 61 ccgctacact tgccagcgcc ctacgcgccg ctcccttctgc tttcttccct tcctttctcg  
 121 ccacggttcgc cggcatcaga ttggctattg gccattgcat acgttgatc catatcataa  
 15 181 tatgtacatt tatattggtt catgtccaac attaccgcca tgttgacatt gattattgac  
 241 tagttattaa tagtaataca ttacggggto attagttoat agcccatata tggagttccg  
 301 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt  
 361 gacgtcaata atgacgtatg ttcccatagt aacgccataa gggactttcc attgacgtca  
 421 atgggtggag tatttacggt aaactgccca ctgggcagta catcaagtgt atcatatgcc  
 481 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgccagta  
 541 catgaacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac  
 601 catggtgatg cgggttttggc agtacctcaa tgggcgtgga tagcggtttg actcacgggg  
 661 atttccaaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg  
 721 ggactttcca aaatgtccta acaactccgc cccattgacg caaatggggg gtaggcgtgt  
 25 781 acggtgggag gtctatataa gcagagctcg tttagtgaaac cgtcagatcg cctggagacg  
 841 ccattccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcgcccg  
 901 ggaacggtgc attggaacgc ggattccccg tgccaagagt gacgtaaata ccgcctatag  
 961 actctatagg cacacccttt tggctcttat gcatgctata ctgttttttg ctggggcct  
 1021 atacaccccc gcttcccttat gctatagtg atggtatagc tttagcctata ggtgtgggtt  
 30 1081 attgaccatt attgaccact cccctattgg tgacgatact ttccattact aatccataac  
 1141 atggctcttt gccacaacta tctctattgg ctatatgcca atactctgtc cttcagagac  
 1201 tgacaacggac tctgtatttt tacaggatgg ggtcccattt attatttaca aattcacata  
 1261 tacaacaacg ccgtcccccgc tgcccgcagt ttttattaaa catagcgttg gatctccacg  
 1321 cgaatctcgg gtacgtgttc cggacatggg ctcttctccg gtacggcgcg agcttccaca  
 35 1381 tccgagccct ggtcccatgc ctccagcgcc tcatggtcgc tccgcagctc cttgctccta  
 1441 acagtggagg ccagacttag gcacagcaca atgccacca ccaccagtgt ccgcacaaag  
 1501 gccgtggcgg tagggatagt gtctgaaaat gagcgtggag attgggctcg caoggtgac  
 1561 gcagatggaa gacttaaggg agcggcagaa gaagatgcag gcagctgagt tgttgtattc  
 1621 tgataagagt cagaggtaac tccggttgcg gtgctgttaa cgggtggagg cagtgtagtc  
 40 1681 tgagcagtac tctgtgtctc cgcgcgcgcc accagacata atagctgaca gactaacaga  
 1741 ctgttccctt ccatgggtct tttctgcagt caccgtcgga ccatgtgcga actcgatatt  
 1801 ttacacgact ctctttacca attctgcccc gaattacact taaaacgact caacagotta  
 1861 acgttggctt gccacgctt acttgactgt aaaaactotca ctcttaccga acttggccgt  
 1921 aacctgcca ccaaacggag aacaaaacat aacatcaaac gaatcgaccg attgttaggt  
 45 1981 aatcgtcacc tccacaaaga gcgactcgt gtataccgtt ggcagctag ctttctctgt  
 2041 tccggcaata cgaatgccat tgaacttgtt gactggtctg atattcgtga gcaaaaacga  
 2101 cttatggtat tgcgagcttc actcgacta cagcgtcgtt ctgttactct ttatgagaaa  
 2161 gcgttcccg cttcagagca atgttcaaaag aaagctcatg accaatctct agccgacctt  
 2221 gcgagcattc taccgagtaa caccacaccg ctcatgttca gtgatgctgg ctttaaagt  
 50 2281 ccatgggtata aatccgttga gaagctgggt tggtaactgt taagtcagat aagaggaaaa  
 2341 gtacaatatg cagacctagg agcggaaaac tggaaaacta tcagcaactt acatgatatg  
 2401 tcatctagtc actcaaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca  
 2461 tgccaaatc tattgtataa atctcgtctt aaaggccgaa aaaatcagcg ctgcacacgg  
 2521 actcattgto accacccgct acotaaaatc tactcagcgt cggcaaaagg gccatggggt  
 55 2581 ctagcaacta acttaoctgt tgaatttcga acacccaaac aacttgtaa tatctattcg  
 2641 aagcgaatgc agattgaaga aaccttccga gacttgaaaa gtccctgccta cggactaggg  
 2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgctgct aatcgccctg  
 2761 atgcttcaac taacatgttg gcttgcgggc gttcatgctc agaaacaagg ttgggacaag  
 2821 cacttccagg ctaacacagt cagaaatoga aacgtactct caacagttcg cttaggcatg  
 60 2881 gaagttttgc ggcattcttg ctacacataa caaagggaag acttactcgt ggctgcaacc  
 2941 ctactagctc aaaatttatt cacacatggg tacgcttttg ggaattatg aggggatccg  
 3001 tctagagcga tccgggatct cgggaaaagc gttggtgacc aaaggtgcct tttatcatca  
 3061 ctttaaaaat aaaaaacaat tactcagtcg ctgttataag cagcaattaa ttatgattga  
 3121 tgccatcatc acaacaaaaa ctgatttaac aaatgggttg tctgccttag aagatatatt  
 65 3181 tgcacattat cttgattata ttattgataa taataaaaaa cttatcccta tccaagaagt  
 3241 gatgcctatc attggttgga atgaactga aaaaaattag ccttgaatac attactggta  
 3301 aggtaaacgc cattgtcagc aaattgatcc aagagaacca acttaagctt ttcctgacgg  
 3361 aatgttaatt ctggttgacc ctgagcactg atgaatcccc taatgatttt ggtaaaaatc

3421 attaaggttaa ggtggatata catcttgtca tatgatcccg gtaatgtgag ttagctcact  
 3481 cattagggcac cccaggcttt acactttatg ctcccgctc gtatgttgtg tggaaattgtg  
 3541 agcggataaac aattttcacac aggaacacagc tatgaccatg attacgccaac gcgcgcaatt  
 5 3601 aaccctcact aaagggaaaca aaagctggag ctccaccgcg gtggcgcccg ctctagaact  
 3661 agtggatccc cggggctgca ggaattcgat atcaagctta tgcataccgc tgacctcgag  
 3721 catcagattg gctattggcc attgcatacg ttgtatccat atcataatat gtacatttat  
 3781 attggtcat gtccaacatt accgcoatgt tgacattgat tattgactag ttattaatag  
 3841 taatcaatta cggggctcatt agttcatagc ccatatatgg agttccgcgt tacataactt  
 3901 acggtaaatg gccgcctcgg ctgaccgccc aacgaccccc gccattgac gtcaataatg  
 10 3961 acgtatgttc ccatagtaac gccaataggg actttccatt gacgtcaatg ggtggagtat  
 4021 ttacggtaaa ctgcccactt ggcagtagat caagtgtatc atatgtcaag tacgccccct  
 4081 attgacgtca atgacggtaa atggcccgcc tggcattatg cccagtagat gaccttatgg  
 4141 gactttcccta ctggcagta catctacgta ttagtcatcg ctattaccat ggtgatgcgg  
 4201 ttttggcagt acatcaatgg gcgtggatag cggtttgact cagggggaatt tccaagtctt  
 15 4261 caccoccatg acgtcaatgg gagtttgttt tggcaccaaa atcaacggga ctttccaaaa  
 4321 tgtcgtaaca actccgcccc attgacgcaa atgggcggtg ggcgtgtacg gtgggagggtc  
 4381 tatataagca gagctcgttt agtgaacogt cagatcgctt ggagacgcca tccacgctgt  
 4441 tttgacctcc atagaagaca cggggaccga tccagcctcc gcggccggga acgggtgcatt  
 4501 ggaacgcgga ttccccgtgc caagagtac gttaagtacc cctatagact ctataggcac  
 20 4561 acccctttgg ctcttatgca tgctatactg tttttggctt ggggcctata ccccccgct  
 4621 tccttatgct ataggtgatg gtatagctta gcctataggt gtgggttatt gaccattatt  
 4681 gaccactccc ctattggtga cgatactttc cattaactaat ccataacatg gctctttgcc  
 4741 acaactatct ctattggcta tatgccaata ctctgtcctt cagagactga cagggactct  
 25 4801 gtatttttac aggatggggt cccatttatt atttacaat tcacatatac aacaacgccc  
 4861 tcccccgctg ccgcagtttt tattaaacat agcgtgggat ctccacgca atctcgggtg  
 4921 cgtgttccgg acatgggctc ttctccggtg gcggcgaggc ttccacatcc gagccctggt  
 4981 cccatgcctc cagcggtcga tggctgcctg gcagctcctt gctcctaaca gtggaggcca  
 5041 gacttaggca cagcacaatg cccaccacca ccagtgtgccc gcacaaggcc gtggcggtag  
 30 5101 ggtatgtgtc tgaaaatgag cgtggagatt gggctcgacac ggcgtgacga gatggaagac  
 5161 ttaaggcagc ggcagaagaa gatgcaggca gctgagttgt tgtattctga taagagtcag  
 5221 aggtaaactcc cgttgcggtg ctgttaacgg tggaggcgag tgtagtctga gcagtagctg  
 5281 ttgctgcccgc gcgcgcaccc agacataata gctgacagac taacagactg ttcccttcca  
 5341 tgggtctttt ctgcagtcac ogtcggatca atcattcacc tctgtacttc ttogtgtgtg  
 35 5401 gtgtttacct atatatctaa atttaatat tctgttatta aaatttaata tatttcgacg  
 5461 atgaatttct caaggatatt ttcttctgtg ttctgttgg ttctggcctt gtcaacagtt  
 5521 tccggtgcgc cagagccgaa aggtaccag gtgcagctgc aggagtcggg gtaaggcttg  
 5581 gtaaaagccg ggggttccct tagagtctcc tgtgcagcct ctggattcac ttccagaaac  
 5641 gccctggatga gctgggtccc ccaggctcca gggaaggggc tggagtgggt cgcccgctatt  
 40 5701 aaaagcaaaa ttgatgggtg gacaacagac tatgtctgac ccgtgaaagg cagattccac  
 5761 atctcaagag atgattcaaa aaacacgtta tatctgcaaa tgaatagcct gaaagccgag  
 5821 gacacagccg tatattactg taccacgggg attatgataa catttggggg agttatccct  
 5881 ccccgcaatt ggggcccagg aaccctggct accgtctcct cagcctccac caagggccca  
 5941 tccgtcttcc cctggcacc ctccccaag agcacctctg ggggcacagc ggccctgggc  
 45 6001 tgcctggta aggactact ccccgaaacg gtgacgggtg cgtggaactc aggcgcctg  
 6061 accagcgccg tgcacacctt tccggctgtc ctacagtcct caggactcta ctctcttagc  
 6121 aacgtgtgtg ccgtgccttc cagcagcttg ggcacccaga cctacatctg caacgtgaat  
 6181 cacaagccca gcaacaccaa ggtggacaag aaagttgagc ccaaatcttg tgacaaaaact  
 6241 cacacatgcc caccgtgccc agcacctgaa ctccctgggg gaccgtcagt ctctctcttc  
 50 6301 ccccaaaaac ccaaggacac cctcatgac tcccggaacc ctgaggtcac atgctgggtg  
 6361 gtggacgtga gccacgaaga cctgaggtc aagttcaact ggtacgtgga cggcgtggag  
 6421 gtgcataatg ccaagacaaa gccgcgggag gagcagtaca acagcacgta ccgtgtggtc  
 6481 agcgtcctca ccgtcctgca ccaggactgg ctgaatggca aggagtacaa gtgcaaggtc  
 6541 tccaacaaag cctcccagc cccatcagag aaaaccatct ccaaagccaa agggcagccc  
 6601 cgagaaccac aggtgtacac cctgccccca tcccgggatg agctgaccaa gaaccaggtc  
 55 6661 agcctgacct gcctggtcaa aggtctctat cccagcgaca tcccggtgga gtgggagagc  
 6721 aatgggcagc cggagaacaa ctacaagacc acgcctcccg tgcctgactc cgacggtccc  
 6781 ttcttctctc acagcaagct caccgtggac aagagcaggt ggcagcaggg gaacctcttc  
 6841 tcatgctccg tgatgcattg ggtctgcac aaccactaca cgcagaagag cctctccctg  
 6901 tctccgggta aagcgccaga gccgaagctt tccatagagc tgacacagcc accctcggtg  
 60 6961 tcaagtgtcc caggacaaac ggcaggatc acctgctctg gagatgcatt gccagaaaaa  
 7021 tatgtttatt ggtaccagca gaagtcaggc caggccccct tgggtgtcat ctatgaggac  
 7081 agcaaacgac cctccgggat cctgagaga ttctctggct ccagctcagg gacaatggcc  
 7141 accttgacta tcagtggggc ccagggtgaa gatgaagggt actactactg ttactcaact  
 7201 gacagcagtg gttatcatag agcgggtgtc agcggaggga ccaagctgac cgtcctaggt  
 65 7261 cagcccaagg ctgccccctc ggtcactctg ttccaccct cctctgagga gcttcaagcc  
 7321 aacaaggcca cactgggtgtg tctcataagt gactcctacc cgggagccgt gacagtggcc  
 7381 tgggaaggcag atagcagccc cgtcaaggcg ggagtggaga ccaccacacc ctccaaacaa  
 7441 agcaacaaca agtacgcggc cagcagctac ctgagcctga cgttgagca gtggaagtcc

	7501	cacaaaagct	acagctgcc	ggctcacgcat	gaagggagca	ccgtggagaa	gacagtggcc
	7561	cctgcagaat	gttcaccgcg	gagggaggga	agggcccttt	ttgaagggg	aggaaacttc
	7621	gcgcatgac	tctctctgtg	ccccccgac	ggaacactga	tgtagagagg	gcoctctgcc
5	7681	attgctgctt	cctctgccct	tctctgtcac	tctgaatgtg	gcttctttgc	tactgccaca
	7741	gcaagaaata	aaatctcaac	atctaaatgg	gtttcctgag	atttttcaag	agtcgttaag
	7801	ccatttcctt	ccccagcacc	ccttgctgca	ggccagtgcc	aggcaccac	ttggctactg
	7861	ctgcccata	gagaaatcca	gttcaatatt	ttccaaagca	aaatggatta	catatgcctt
	7921	agatcctgat	taacaggigt	tttgtattat	ctgtgctttc	gcttcaccca	cattatccca
	7981	ttgcctcccc	tcgagggggg	gcccggtagc	caattcgccc	tatagttagt	cgtattacgc
10	8041	gcgctcaactg	gcgctcggtt	tacaacgtcg	tgactgggaa	aacctggcg	ttaccoaact
	8101	taatgcgctt	gcagcacatc	cccccttgc	cagctggcgt	aatagcgaag	aggcccgac
	8161	cgatcgccct	ttcccaacagt	tgccgagcct	gaatggcgaa	tggaatgtt	aagcgttaat
	8221	attttgttaa	aattcgcggt	aaatttttgt	taaatcagct	cattttttta	ccaataggcc
	8281	gaaatcgcca	aaatccctta	taaatcaaaa	gaatagaccg	agatagggtt	gagtggtgtt
15	8341	ccagtgttga	acaagagtcc	actattaaag	aacgtggact	ccaacgtcaa	agggcgaaaa
	8401	accgtctatc	agggcgatgg	cccactactc	cgggatcata	tgacaagatg	tgtatccacc
	8461	ttaacctaat	gattttttac	aaaatcatta	ggggattcat	cagtgtctag	ggtcaacgag
	8521	aatataacatt	ccgtcagga	agcttatgat	gatgatgtgc	ttaaaaactt	actcaatggc
	8581	tggttatgca	tatcgcaata	catgcaaaaa	acctaaaaaga	gcttgccgat	aaaaaaggcc
20	8641	aatttattgc	tattttaccgc	ggctttttat	tgagcttgaa	agataaataa	aatagatagg
	8701	ttttatttga	agctaaatct	tctttatcgt	aaaaaatgcc	ctcttgggtt	ttcaagagg
	8761	tcattatatt	tcgagggaata	acatcatttg	gtgacgaat	aactaagcac	ttgtctcctg
	8821	tttactcccc	tgagcttgag	gggttaacat	gaaggtcatc	gatagcagga	taataatata
25	8881	gtaaacgct	aaaccaataa	ttccaaatcca	gccatcccaa	atttgtagtg	aatgattata
	8941	aataacagca	aacagtaagt	ggccaataac	accgggttga	ttgtaaggc	tcaccaataa
	9001	ttccctgtaa	gcaccttgct	gatgactctt	tgtttgata	gacatcactc	cctgtaatgc
	9061	aggttaagcg	atcccaccac	cagccaataa	aattaaaaa	gggaaaaact	accaaccttc
	9121	agatataaac	gctaaaaagg	caaatgcact	actactctga	ataaatccga	gcagtactgc
	9181	cgttttttgc	cccatttagt	ggctattctt	cctgcccaca	aggcttgga	tactgagtgt
30	9241	aaaagacca	gaccgcgtat	gaaaagccaa	ccatcatgct	attcatcatc	acgatttctg
	9301	taatagcacc	acaccgtgct	ggattggcta	tcaatgcgct	gaaataataa	tcaacaaatg
	9361	gcacogttta	ataagtgatg	tatacogato	agcttttggt	cccttagtg	aggggttaatt
	9421	gcgctgttg	cgtaatcatg	gtcatagctg	tttctgtgtg	gaaattgtta	ttccgtcaca
	9481	attccacaca	acatacagagc	cggaagcata	aaagtgtaaag	cctggggtgc	ctaagttagt
35	9541	agctaactca	catttaattgc	gttgccgtca	ctgcccgtt	tccagtcggg	aaacctgtcg
	9601	tgccagctgc	attaatgaat	cggccaaacg	gcggggagag	gcggtttgct	tattggcgcc
	9661	tcttcgctt	cctcgctcac	tgactcgctg	cgctcggtcg	ttcggtcgcg	gcgagcggtg
	9721	tcagctcaact	caaaggcggt	aatcacggta	ttccacagaat	caggggataa	cgcaggaaag
40	9781	aacatgtgag	caaaaggccc	gcaaaaggcc	aggaaccgta	aaaaggccgc	gttgctggcg
	9841	tttttccata	ggctccgccc	ccctgacgag	catcacaaaa	atcgacgctc	aagtacaggg
	9901	tgccgaaacc	cgacaggact	ataagatac	caggcggttc	cccttggaag	ctccctcggt
	9961	cgtctctctg	ttccgaccct	gcccgttacc	ggatacctgt	ccgcctttct	cccttcggga
	10021	agcgtggcgc	tttctcatag	ctcacgctgt	aggtatctca	gttcggtgta	ggtcgttcgc
45	10081	ttcaaagctgg	gctgtgtgca	cgaaccccc	gttcagcccg	accgctgcgc	cttatccggt
	10141	aactatcgtc	ttgagtcocaa	cccggtaaga	cacgacttat	cgccactggc	agcagccact
	10201	ggtaacagga	ttagcagagc	gaggtatgta	ggcgggtgta	cagagttctt	gaagtgggtg
	10261	ccataactacg	gtacacatg	aaggacagta	tttggtatct	gcgctctgct	gaagccagtt
	10321	accttcggaa	aaagagtgtg	tagctcttga	ttccggcaaac	aaaccaccgc	tggttagcgtt
	10381	ggtttttttg	tttgcaagca	gcagattacg	cgcagaaaaa	aaggatctca	agaagatcct
50	10441	ttgatctttt	ctacgggggtc	tgacgctcag	tggaacgaaa	actcacgtta	agggattttg
	10501	gtcatgagat	tatcaaaaag	gatcttcacc	tagatccttt	taaattaaaa	atgaagtttt
	10561	aaatcaatct	aaagtatata	tgagtaaaact	tggtctgaca	gttaaccaatg	cttaatcagt
	10621	gaggcaccta	tctcagcgat	ctgtctattt	cgttcaccca	tagttgcttg	actcccgcgc
55	10681	gtgtagataa	ctacagatac	ggaggggtta	ccatctgggc	ccagtgtgc	aatgataccg
	10741	cgagaccac	gctcacccgc	ttccagattta	tcagcaataa	accagccagc	cggaaggggc
	10801	gagcgcagaa	gtggctcctg	aaactttatcc	gcctccatcc	agtcctattaa	ttgttgccgg
	10861	gaagctagag	taagtgttct	gccagttaat	agtttgccga	acgttggtgc	cattgctaca
	10921	ggcatcgtgg	tgtaacgctc	gtcgttttgt	atggcttcat	tcagctccgg	ttcccaacga
	10981	tcaaggcgag	ttacatgac	ccccatgttg	tgcaaaaaag	cgttagctc	cttcggctct
60	11041	ccgatcgttg	tcagaagtaa	gttgcccgca	gtgttatcac	tcagtgttat	ggcagcactg
	11101	cataattctc	ttactgtcat	gccatccgta	agatgctttt	ctgtgactgg	tgagtactca
	11161	accaagtcat	tctgagaata	gtgtatgcgg	cgaccgagtt	gctcttgccc	ggcgtcaata
	11221	cgggataata	ccgcgccaca	tagcagaact	ttaaaagtgc	tcacatctgg	aaaacgctct
	11281	ctggggcgaa	aactctcaag	gatcttaccg	ctgttgagat	ccagttcgat	gtaacccact
65	11341	cgtgcaccca	actgatcttc	agcatctttt	actttcacca	gcgtttctgg	gtgagcaaaa
	11401	acaggaaggc	aaaatgccgc	aaaaaaggga	ataaggcgca	cacggaaatg	ttgaataact
	11461	atactcttcc	tttttcaata	ttattgaagc	atattacagg	gttattgtct	catgagcgga
	11521	tacatatattg	aatgtattta	gaaaaataaa	caaatagggg	ttccgcgcac	atttccccga

11581 aaagtgccac

SEQ ID NO: 49 pTnMCS (Chicken OVep+OVg'+ENT+proins+syn polyA)

5 1 ctgaogcgcg ctgtagcggc gcattaagcg cggcggtgt ggtggttacg cgcagcgtga  
61 ccgctacact tggcagcgcc ctagcgcccg ctcccttcgc tttcttccct tcctttctcg  
121 ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgtatc catatcataa  
181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac  
241 tagttattaa tagtaataa ttacggggtc attagtcat agcccatata tggagttccg  
10 301 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgccatt  
361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca  
421 atgggtggag tatttacggt aaactgcccc cttggcagta catcaagtgt atcatatgcc  
481 aagtacggcc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta  
541 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac  
15 601 catggtgatg cggttttggc agtacatcaa tggcggtgga tagcggtttg actcacgggg  
661 atttccaaat ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg  
721 ggactttcca aaatgtcgta acaactccgc cccattgaag caaatggcg gtaggcgtgt  
781 acgggtggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg  
841 ccatccacgc tgttttgacc tccatagaag acacggggac cgtaccagcc tccgcgcccg  
20 901 ggaacggtgc attggaacgc ggattccccg tgccaagagt gacgtaagta cgcctatag  
961 actctatagg cacaccctt tggctcttat gcatgotata ctgtttttgg cttggggcct  
1021 atacaccccc gcttctcttat gctataggtg atggtatagc ttgacctata ggtgtgggtt  
1081 attgaccatt attgaccact cccctatttg tgacgatact ttccattact aatccataac  
1141 atggctcttt gccacaacta tctctatttg ctatatgcca atactctgtc cttcagagac  
25 1201 tgacacggac tctgtatttt tacaggatgg ggtcccattt atattttaca aattcacata  
1261 tacaacaacg ccgtcccccg tgcgcgagc ttttattaaa catagcgtgg gatctccacg  
1321 cgaatctcgg gtacgtgttc cggacatggg ctcttctccg gtacggcggg agcttccaca  
1381 tccgagccct ggtcccatgc ctccagcgcc tcatggtcgc tccgagctc cttgctcta  
1441 acagtgagg ccagacttag ccacagcaca atgcccacca ccaccagtgt gccgcacaag  
30 1501 gcogtggcgg tagggtatgt gtctgaaat gagcgtggag attgggctcg cacggtgac  
1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgtatc  
1621 tgataagagt cagaggtaac tccggttgog gtgctgttaa cggtgaggg cagtgtagtc  
1681 tgagcagtac tcggtgtctgc cgcgcgcgcc accagacata atagctgaca gactaacaga  
1741 ctgttctctt ccatgggtct tttctgcagt caccgtcgga ccatgtcgga actcgatatt  
35 1801 ttacacgact ctctttacca attctgcccc gaattacact taaaacgact caacagctta  
1861 acgtttggctt gccacgcat ctctgactgt aaaactctca ctcttacga acttggccgt  
1921 aacctgccaa ccaaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt  
1981 aatcgtcacc tccacaaaga gcgactcgct gtataccgtt ggcagctag ctttatctgt  
40 2041 tccggcaata cgtatcccat tgtactgtt gactggtctg atattcgtga gcaaaaaoga  
2101 ctataggtat tgcgagcttc agtcgcaata cacggtcggt ctgttactct ttatgagaaa  
2161 gcgttcccg cttcagagca atgttcaaag aaagctcatg accaatttct agccgacct  
2221 gcgagcatte taccgagtaa caccacaccg ctcatgttca gtgatgctgg ctttaaagt  
2281 ccatggtata aatccgttga gaagctgggt tggactcggt taagtcgagt aagaggaaaa  
45 2341 gtacaatatg cagacctagg agcggaaaac tggaaaacta tcagcaactt acatgatag  
2401 tcactatagtc actcaaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca  
2461 tgccaaatct tattgtataa atctcgctct aaaggccgaa aaaatcagcg ctccgacagg  
2521 actcattgtc accaccogtc actaaaatc tactcagcgt cggcaaaagg gccatgggtt  
2581 ctagcaacta acttacctgt tgaattoga acacccaaac aacttgttaa tatctattcg  
2641 aagcgaatgc agattgaaga aaccttccga gacttgaaaa gtocctgcta cggactagge  
50 2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgctgct aatcgccctg  
2761 atgcttcaac taacatgttg gcttgcgggc gttcatgctc agaaacaagg ttgggacaag  
2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg  
2881 gaagttttgc ggcattcttg ctacacaata acaagggaag acttactcgt ggtgcaacc  
2941 ctactagctc aaaaatttatt cacacatggt taogcttttg ggaaattatg aggggacatc  
55 3001 tctagagcga tccgggatct cgggaaaagc gttggtgacc aaaggtgctt tttatcatca  
3061 ctttaaaaaa aaaaaacaat tactcagtc ctgttataag cagcaattaa ttatgattga  
3121 tgccatcatc acaacaaaaa ctgatttaac aaatgggttg tctgctttag aaagtatat  
3181 tgaacattat cttgattata ttattgataa taataaaaaa cttatcccta tccaagaagt  
3241 gatgcctatc attggttgga atgaacttga aaaaaattag ccttgaatac attactggta  
60 3301 aggtaaacgc cattgtcagc aaattgatcc aagagaacca acttaagct ttctgacgg  
3361 aatgttaatt ctogttgacc ctgagcactg atgaatcccc taatgatttt ggtaaaaaatc  
3421 attaagttaa ggtggataca catcttgtca tatgatcccg gtaatgtgag tttagctact  
3481 cattagcac cccaggtctt acactttatg cttccggctc gtatgtgtg tggaaattgt  
3541 agcggataac aatttcacac aggaacacgc tatgacctg attacgcaa gcgcgcaatt  
65 3601 aaccctcact aaagggaaca aaagctggag ctccaccgag gtggcgccg cctctagaact  
3661 agtggatccc ccgggctgca gaaaaatgcc aggtggacta tgaactcaca tccaaaggag  
3721 cttgacctga tacctgattt tcttcaaac ggggaaacaa cacaatccca caaaacagct  
3781 cagagagaaa ccatcactga tggctacagc accaaggtat gcaatggcaa tccattcgac

	3841	attcatctgt	gacctgagca	aatgatttta	tctctccatg	aatggttgct	tctttccctc
	3901	atgaaaaggc	aatttccaca	ctcacaaat	gcaacaaaga	caaacagaga	acaattaatg
	3961	tgctccttcc	taatgtcaaa	attgtagtgg	caaagaggag	aacaaaatct	caagttctga
5	4021	gtagggttta	gtgattggat	aagaggcttt	gacctgtgag	ctcacctgga	cttcatatcc
	4081	ttttggataa	aaagtgtctt	tataactttc	aggtctccga	gtctttatcc	atgagactgt
	4141	tggttttagg	acagaccac	aatgaaatgc	ctggcatagg	aaagggcagc	agagccttag
	4201	ctgacctttt	cttgggacaa	gcattgtcaa	acaatgtgtg	acaaaactat	ttgtactgct
	4261	ttgcacagct	gtgctgggca	gggcaatcca	ttgccacctt	tcccaggtaa	ccttccaaat
10	4321	gcaagaagat	tggtgtctac	tctctctaga	aagcttctgc	agactgacat	gcatttccata
	4381	ggtagagata	acattttactg	ggaagcacat	ctatcatcat	aaaaagcagg	caagattttc
	4441	agactttctt	agtggctgaa	atagaagcaa	aagacgtgat	taaaaacaaa	atgaaacaaa
	4501	aaaaatcagt	tgatacctgt	ggtgtagaca	tccagcaaaa	aaatatttatt	tgcaactacca
	4561	tcttgtctta	agtcctcaga	cttggcaagg	agaatgtaga	tttctacagt	atatatgttt
	4621	tcacaaaagg	aaggagagaa	acaaaagaaa	atggcactga	ctaaacttca	gctagtggta
15	4681	taggaaagta	attctgtcta	acagagattg	cagtgtatct	tatgtatgtc	ctgaagaatt
	4741	atgttgtact	tttttccccc	atttttaaat	caaacagctg	tttacagagg	tcagaatggt
	4801	ttcttttact	tttgtcaatt	ctattatttc	aatacagaac	aatagcttct	ataactgaaa
	4861	tatatgtgct	attgtataat	atgattgtcc	ctcgaacctc	gaacactcct	ccagctgaat
	4921	ttcacaaatc	ctctgtcatc	tgccaggcca	ttaagttatt	catggaagat	ccttgaggaa
20	4981	cactgcaagt	tcatactcata	aacacatttg	aaattgagta	ttgttttgca	ttgtattggg
	5041	ctatgttttg	ctgtatcctc	agaaaaaaag	tttgttataa	agcattccaca	cccataaaaa
	5101	gatagattta	aattattccag	ctataggaaa	gaaagtgcgt	ctgctcttca	ctctagtctc
	5161	agttggctcc	ttcacatgca	tgcttcttta	tttctcctat	tttgtcaaga	aaataatagg
	5221	tcacgtcttg	ttctcactta	tgctctgcct	agcatggctc	agatgcacgt	tgtagataca
25	5281	agaagatcca	aatgaaacag	actctctgtc	tggtactaca	accatagtaa	taagcaactc
	5341	aactaataat	tgctaattat	gttttccatc	tctaagggtc	ccacattttt	ctgttttctt
	5401	aaagatccca	ttatctgtgt	gtaactgaag	ctcaatggaa	catgagcaat	atttccagct
	5461	cttctctccc	atccaacagt	cctgatggat	tagcagaaca	ggcagaaaaa	acattgttac
	5521	caagaattaa	aaactaatat	ttgctctcca	ttcaatccaa	aatggacctc	ttgaaactaa
30	5581	aatctaacc	aatcccatta	aatgatttct	atggcgctca	aggtcaaaact	tctgaaggga
	5641	acctgtgggt	gggtcacat	tcaggctata	tattcccccag	ggctcagcca	gtggatcaac
	5701	atcacgctag	aaagctgtat	tgctttagc	actcaagctc	aaaagacaac	tcagagttca
	5761	ccatgggctc	catcggcgca	gcaagcatgg	aattttgttt	tgatgtattc	aaggagctca
	5821	aatgtccacca	tgccaatgag	aacatcttct	actgccccat	tgccatcatg	tcagctctag
35	5881	ccatgggtata	cctgggtgca	aaagacagca	ccaggacaca	gataaataag	gttgttcgct
	5941	ttgataaaat	tcaggagttc	ggagacagta	ttgaagctca	gtgtggcaca	ctgttaaacg
	6001	ttcactcttc	acttagagac	atcctcaacc	aaatcaccac	accaaatgat	gtttattcgt
	6061	tcagccttgc	cagtacagct	tatgctgaag	agagataccc	aatcctgcca	gaataactgc
	6121	agtgtgtgaa	ggaactgtat	agaggaggct	tggaacctat	caactttcaa	acagctgcag
40	6181	atcaagccag	agagctcatc	aattcctggg	tagaaagtca	gacaaatgga	attatcagaa
	6241	atgtccttca	gccaagctcc	gtggattctc	aaactgcaat	ggttctgggt	aatgccattg
	6301	tcttcaaaag	actgtgggag	aaaacattta	aggatgaaga	cacacaagca	atgcctttca
	6361	gaagtactga	gcaagaaaag	aaacctgtgc	agatgatgta	ccagattggt	ttatttagag
	6421	tggtcatcaat	ggcttctgag	aaaatgaaga	tccctggagct	tccatttgcc	agtgggacaa
45	6481	tgagcatggt	gggtctgttg	cctgatgaag	tctcaggcct	tgagcagctt	gagagtataa
	6541	tcaactttga	aaaactgact	gaatggacca	gttctaattg	tatggaagag	aggaagatca
	6601	aaagtgaact	acctcgcatg	aaagtggagg	aaaaatacaa	cctcacatct	gtcttaattg
	6661	ctatgggcat	tactgacgtg	tttagctott	cagccaatct	gtctggcatc	tcctcagcag
	6721	agagcctgaa	gatatctcaa	gctgtccatg	cagcacatgc	agaaatcaat	gaagcaggca
50	6781	gagaggtggt	agggctcagca	gaggctggag	tggaatgctg	aagcgtctct	gaagaattta
	6841	gggctgacca	tccattcctc	ttctgtatca	agcacatcgc	aaccaacgcc	gttctcttct
	6901	ttggcagatg	tgtttctccg	cggccagcag	atgacgcacc	agcagatgac	gcaccagcag
	6961	atgacgcacc	agcagatgac	gcaccagcag	atgacgcacc	agcagatgac	gcaacaacat
	7021	gtatcctgaa	aggctcttgt	ggctggatcg	gcctgctgga	tgacgatgac	aaatttgtga
55	7081	accaaacacct	gtgcggctca	cacctgggtg	aagctctcta	cctagtgtgc	ggggaacgag
	7141	gcttctctca	cacacccaag	acccgcccgg	aggcagagga	cctgcagggtg	gggcagggtg
	7201	agctggggcg	gggcccgtgt	gcaggcagcc	tgacgcccct	ggccctggag	gggtcccctg
	7261	agaagcgtgg	cattgtggaa	caatgctgta	ccagcatctg	ctcccctctc	cagctggaga
	7321	actactgcaa	ctaggcgccc	taaagggcga	attatgcggg	ccgctctaga	ccaggcgccct
60	7381	ggatccagat	cacttctggc	taataaaaga	tcagagctct	agagatctgt	gtgttggttt
	7441	tttgtggatc	tgctgtgcct	tctagtggcc	agccatctgt	tgtttgcccc	tccccgtgc
	7501	cttccctgac	cctggaaggt	gccactccca	ctgtcctttc	ctaataaaat	gaggaaattg
	7561	catcgcatgt	tctgagttag	tgctatttca	ttctgggggg	tggggtgggg	cagcacagca
	7621	agggggagga	ttgggaagac	aatagcaggg	atgctgggga	tgcggtgggc	tctatgggta
65	7681	ctctctctct	tctctctctc	tctctctctc	tctctctctc	tcggtaacct	tctcgagggg
	7741	gggcccggta	cccaattcgc	cctatagtga	gtcgtattac	gcgcgctcac	tgcccgctcg
	7801	tttacaacgt	cgtgactggg	aaaaccctgg	cgttaaccac	cttaatcgcc	ttgcagcaca
	7861	tccccctttc	gccagctggc	gtaatagcga	agaggcccg	accgatcgcc	cttcccaaca

	7921	gttgcgcagc	ctgaatggcg	aatggaaatt	gtaagcgta	atattttgtt	aaaattcgcg
	7981	ttaaattttt	gttaaatcag	ctcatttttt	aaccaatagg	cgaatcgg	caaaatccct
	8041	tataaatcaa	aagaatagac	cgagataggg	ttgagtggtg	ttccagtttg	gaacaagagt
5	8101	ccactattaa	agaacgtgga	ctccaacgtc	aaaggcgaa	aaacggtcta	tcaggcgat
	8161	ggcccaactac	tccgggatca	tatgacaaga	tgtgtatcca	ccttaactta	atgattttta
	8221	ccaaaatcat	taggggatcc	atcagtgctc	agggtcaacg	agaattaaca	ttccgtcagg
	8281	aaagcttatg	atgatgatgt	gcttaaaaaa	ttactcaatg	gctgggtatg	catatcgcaa
	8341	tacatgcgaa	aaacctaaaa	gagcttgccg	ataaaaaagg	ccaattttat	gctatttacc
10	8401	gcggcttttt	attgagcttg	aaagataaat	aaaatagata	ggtttttatt	gaagctaaat
	8461	cttcctttatc	gtaaaaaatg	ccctcttggt	ttatcaagag	ggtcattata	tttcgcgaa
	8521	taacatcatt	tggtagcgaa	ataactaagc	acttgtctcc	tgtttactcc	cctgagcttg
	8581	aggggttaac	atgaaggtca	tcgatagcag	gataataata	cagtaaaacg	ctaaaccaat
	8641	aatccaaatc	cagccatccc	aaattggtag	tgaatgatta	taataaacag	caaacagtaa
	8701	tggggccaata	acaccggttg	cattggtaag	gctcaccat	aatccctgta	aagcaccttg
15	8761	ctgatgactc	tttgtttgga	tagacatcac	tcctgtaat	gcaggtaaa	cgatccacc
	8821	accagccaat	aaaattaaaa	cagggaacac	taaccaacct	tcagatataa	acgctaaaaa
	8881	ggcaaatgca	ctactatctg	caataaatcc	gagcagtact	gccgtttttt	cgccatttta
	8941	gtggcttattc	ttcctgcccac	aaaggcttg	aatactgagt	gtaaaagacc	aagaccctga
20	9001	atgaaaagcc	aaccatcatg	ctatcoatca	tcacgatttc	tgtaatagca	ccacaccgtg
	9061	ctggattggc	tatcaatgcg	ctgaataaat	aatcaacaaa	tggtatcggt	aaataagtga
	9121	tgtataccga	tcagcttttg	ttcccttttag	tgagggttaa	ttgcgcgctt	ggcgtaattca
	9181	tgggtcatagc	tgtttcctgt	gtgaaattgt	tatccgctca	caattccaca	caacatacga
	9241	gccggaagca	taaagtgtaa	agcctggggt	gctaataag	tgagctaact	cacatttaatt
25	9301	gcgttgcgct	cactgcccgc	tttccagtcg	ggaaacctgt	cgtgcccagc	gcattaatga
	9361	atcgcgccaac	gcgcggggag	aggcggtttg	cgtattgggc	gctcttcgc	ttcctcgctc
	9421	actgactcgc	tgcgctcggt	cggtcggctg	cgcgagcgg	tatcagctca	ctcaaaaggc
	9481	gtaatacgg	tatccacaga	atcaggggat	aacgcaggaa	agaacatgtg	agcaaaaggc
	9541	cagcaaaagg	ccaggaaacc	taaaaaggcc	gcgttgctgg	cgtttttcca	taggctccgc
30	9601	ccccctgacg	agcatcacia	aaatcgacgc	tcaagtcaga	gggtggcgaaa	ccgcacagga
	9661	ctataaagat	accaggcggt	ttccccttga	agctccctcg	tgcgctctcc	tgttccgacc
	9721	ctgcgcgtta	cgggatacct	gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcat
	9781	agotcacgct	gtagggtatc	cagttcggtg	taggtcgctc	gctcaagct	gggctgtgtg
	9841	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	gtaactatcg	tcttgagtc
35	9901	aaccccggtaa	gacacgactt	atcgccactg	gcagcagcca	ctggtaacag	gatttagcaga
	9961	gcgaggtatg	tagggcggtg	tacagagttc	ttgaagtgg	ggcctaacta	cggtacact
	10021	agaaggacag	tatttgggat	ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt
	10081	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	gtggtttttt	tgtttgcaag
	10141	cagcagatta	cgcgcagaaa	aaaaggatct	caagaagatc	ctttgatctt	ttctacgggg
40	10201	tctgacgctc	agtggaacga	aaactcaoct	taagggtatt	tggtcatgag	attatcaaaa
	10261	aggatcttca	cctagatcct	tttaaatata	aatgaagtt	ttaaatcaat	ctaaagtata
	10321	tatgagtaaa	cttgggtctga	cagttaccaa	tgcttaatca	gtgaggcacc	tatctcagcg
	10381	atctgtctat	ttcgttcac	catagttgcc	tgaactcccc	tcgtgtagat	aactacgata
	10441	cgggagggct	tacatctgg	ccccagtgct	gcaatgatac	cgcgagacc	acgctcaccg
45	10501	gctccagatt	tatcagcaat	aaaccagcca	gcccgaagg	ccgagcgag	aagtggctct
	10561	gcaactttat	ccgcctccat	ccagctctatt	aatgtgtgcc	gggaagctag	agtaagtgt
	10621	tcgcaggtta	atagtttg	caacgittgt	gccattgcta	caggcatcgt	ggtgtcacgc
	10681	tcgtcgtttg	gtatggcttc	attcagctcc	ggttcccaac	gatcaaggcg	agttacatga
	10741	ttcccatgt	tgtgcaaaaa	agcggttagc	tccttcggtc	ctccgatcgt	tgtcagaagt
50	10801	aagttggcgc	cagttttatc	actcatggtt	atggcagcac	tgcataattc	tcttactgtc
	10861	atgccatccg	taagatgctt	ttctgtgact	ggtgagtact	caaccaagtc	attctgagaa
	10921	tagtgtatgc	ggcgaccgag	ttgctcttgc	ccggcgctca	tacgggataa	taccgcgcca
	10981	catagcagaa	ctttaaaagt	gctcatcatt	ggaaaacgtt	cttcggggcg	aaaactctca
	11041	aggatcttac	cgctgttgag	atccagttcg	atgtaaccca	ctcgtgcacc	caactgatct
55	11101	tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	aaacaggaag	gcaaaatgcc
	11161	gcaaaaaagg	gaataaggcg	gacacggaaa	tgttgaatac	tcatactctt	cctttttcaa
	11221	tattattgaa	gcatttatca	gggttattgt	ctcatgagcg	gatacatatt	tgaatgtatt
	11281	tagaaaaata	aacaaatagg	ggttccgcgc	acatttcccc	gaaaagtgcc	ac
60							